

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 08:56:12 ; Search time 34.94 Seconds
(without alignments)
2384.240 Million cell updates/sec

Title: US-09-810-186-1
Perfect score: 3996
Sequence: 1 MTWTPKISVNDGKLTVHGT.....TFSIPVSEEMRMSIEIQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_033802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3996	100.0	750	20	AA17418 Soybean raffinose
2	2979.5	74.6	751	21	AA770980 Wheat raffinose sy
3	2874	71.9	756	21	AA770975 Corn raffinose syn
4	2827.5	70.8	763	21	AA770977 Rice raffinose syn
5	2117.5	53.0	770	21	AA770976 Rice raffinose syn
6	2110.5	52.8	841	21	AA770981 Wheat raffinose sy
7	1590	39.8	508	21	AA770974 Corn raffinose syn
8	1430.5	35.8	758	21	AA770978 Soybean raffinose
9	1415.5	35.4	777	20	AA732074 Mustard raffinose
10	1394	34.9	784	19	AAW53570 Cucurbit raffinose
11	1394	34.9	784	20	AA17417 Cucurbit raffinose

12	1385	34.7	781	19	AAW57887
13	1385	34.7	781	20	AAV30143
14	1385	34.7	781	22	AAW49400
15	1379	34.5	783	20	AAV32073
16	1370.5	34.3	799	19	AAW57886
17	1370.5	34.3	799	20	AAV30142
18	1367.5	34.2	780	22	AAW57889
19	1117.5	28.0	587	19	AAW57888
20	1117.5	28.0	587	20	AAV30144
21	1095.5	27.4	572	20	AAV32075
22	538.5	13.5	229	21	AAW28628
23	484.5	12.1	265	20	AAV32072
24	477	11.9	271	19	AAW57889
25	477	11.9	271	20	AAV30145
26	376.5	9.4	193	21	AAW28629
27	235.5	5.9	120	21	AAW28630
28	178	4.5	45	21	AAV70979
29	178	4.5	87	21	AAW15267
30	178	4.5	157	21	AAW15266
31	120.5	3.0	2122	22	ABG22590
32	120.5	3.0	2652	22	ABG11747
33	120.5	3.0	2657	22	ABG04691
34	119.5	3.0	920	22	AAV97652
35	119.5	3.0	1074	19	AAW64186
36	117.5	2.9	4544	15	AAW47861
37	117.5	2.9	4544	15	AAW60517
38	117.5	2.9	4544	22	AAW79091
39	117.5	2.9	4563	22	ABW13533
40	116	2.9	1732	22	ABW71861
41	115	2.9	478	19	AAW64173
42	115	2.9	914	20	AAW86429
43	115	2.9	928	20	AAW88418
44	111	2.8	2209	3	AAW20037
45	109.5	2.7	780	15	AAW62487

ALIGNMENTS

RESULT	1
AA17418	
ID	AA17418 standard; Protein: 750 AA.
XX	
AC	AA17418;
XX	
DE	29-JUL-1999 (first entry)
XX	
DE	Soybean raffinose synthase.
XX	
KW	Raffinose synthase; sucrose; galactinol.
XX	
OS	Glycine max.
XX	
EN	JP11123080-A.
XX	
PD	11-MAY-1999.
XX	
PE	24-OCT-1997; 97JP-0292969.
XX	
PR	24-OCT-1997; 97JP-0292969.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
DR	WPI: 1999-340516/29.
XX	
DR	N-PSDB: AAW61239.
XX	
PT	New raffinose synthase gene - for production of raffinose from
XX	sucrose and galactinol
PS	Claim 2; Page 32-34; 37pp: Japanese.
XX	
CC	The present invention describes a raffinose synthase, having an actively
CC	of forming raffinose from sucrose and galactinol. The raffinose synthase

Soybean raffinose
Amino acid sequenc
Soybean raffinose
Soybean raffinose
Broad bean raffin
Amino acid sequenc
Soybean protein: S
Japanese artichoke
Amino acid sequenc
Repeased raffinose
Arabidopsis thalia
Soybean raffinose
Corn raffinose syn
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Soybean raffinose
Arabidopsis thalia
Arabidopsis thalia
Novel human diagno
Novel human diagno
Novel human diagno
Pit-1extraRSTM/cy
Lettuce resistance
Alpha 2-Macroglobu
Human alpha-2-MK.
Human protein SEQ
Human LDL receptor
Drosophila melanog
Lettuce resistance
Chlamydia pneumoni
Sequence encoded b
Truncated FLT SV68

CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from soybean.

XX Sequence 750 AA;

Query Match 100.0%; Score 3996; DB 20; Length 750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTPVPRKISVNDGKLVHGKTLTGVPNDVNIETPGSGRGIVTGAFVGATASHSKSLHVFPM 60
DB 1 mtpvprkisvndgklvngkltlgvpndvnltpgsrgivtgafvgatashskslhvfpm 60
OY 61 GVLGSLRPMCCFRPKLMMTORMGTGRDVPLETOFMLESKSEETDGENSPIITYVL 120
DB 61 gvlgsrlpmccfrpkllmmtrmgctgrdvpлетofmleskesetdgenSPIITYVL 120
OY 121 LLEGQFRAVLQGNDRKNEIETCLESQDNAAVETDQGLHMYVMHAGNPREVINQAVKAVEKH 180
DB 121 llegqfrevlqgnrkneietclesqdnaveTdqglhmymhagnprevinqavkavekh 180
OY 181 MOTFLHREKRLPSCLDFGMCWTDAFYTDVTAEGVEEGKSLSQGGTPRFLIIDGMQ 240
DB 181 moflhrckrlpscldfgmcwtadafytdvtaegveegkslsqggtpRFLIIDGMQ 240
OY 241 QIENKAKDATECLVOEGAOFATRLTGIKENTKFOKLIQNDQNSGLKHLVHGAKQHNHV 300
DB 241 qienkakdateclvgegaofatrltgikentkfokliqndqnsGLKHLVHGAKQHNHV 300
OY 301 NVYVNHALAGYWGKVPRAATGMEHVDLAVPYOSPGVLGNOPRIYMDSLAVHGLGLVHP 360
DB 301 nvvvnhalagywgkvpaaTgmenyvdlaaypvgspgvlgngpdiymdsleVhglglvhp 360
OY 361 KVFENYNEHLAYLASCGVDGKVDVQNIETLGAGHGRVSLRSYHHALEASIASNFT 420
DB 361 kvfenynehlaylasCGVDGKVDVQNIETLGAGHGRVSLRSYHHALEASIASNFT 420
OY 421 DNGCIACMCHNTDGLYSAKQAIYASDDFPRDPASHTIHSSVAYNSLLEGFEMQPDW 480
DB 421 dngciacmchntdglYsAkQAIYASDDFPRDPASHTIHSSVAYNSLLEGFEMQPDW 480
OY 481 DMFHSHPAADYHAARAIIGCCPIYVSKPGNHNFDLKTIVLPDGSYLRAQLPGRPTRD 540
DB 481 dmfhsHPAADYhaaraIIGCCPIYVSKPGNHNFDLKTIVLPDGSYLRAQLPGRPTRD 540
OY 541 SLFVDPARDRTSLIKIWMLNKCSGVVGVFNCGAGWCKIEKTRIHDTSPGTLASVCAS 600
DB 541 slfvdPARDRTSLIKIwmLNKCSGVVGVFncgagwckIEKTRIHDTSPGTLASVCAS 600
OY 601 DVDLITOVAGAMIDTYYAARSGEVRILPKVGSIPITLKLFEHLHPFOIOIASI 660
DB 601 dvdLITOVAGAMIDTYYAARSGEVRILPKVGSIPITLKLFEHLHPFOIOIASI 660
OY 661 SFAAIGLDMFTMGAVEQVEIHNRATKTIALSVKGRGREGVVSORBLKCVVGAETD 720
DB 661 sfaaigldmftmgaveqveIHNRATKTIALSVKGRGREGVVSORBLKCVVGAETD 720
OY 721 FNTDSETGLTFSIPYSPEEMTRMSLEIOV 750
DB 721 fntdsetglTFSIPYSPEEMTRMSLEIOV 750

RESULT 2

AAT70980 AAY70980 standard; Protein; 751 AA.

XX AAY70980;

XX 09-AUG-2000 (first entry)

DE Wheat raffinose synthase from clone wlm24.pk0021.h1.
XX Wheat; raffinose synthase; raffinose saccharide; soybean;
KW clone wlm24.pk0021.h1; nutritional; soy protein.

XX Triticum aestivum.

PN WO200024915-A2.

XX 04-MAY-2000.

PD 22-OCT-1999; 99WO-US24923.

XX 23-OCT-1998; 98US-0105451.

XX (DUPO) DU POINT DE MEMOIRS & CO E. I.

PA Allen SM, Hltz WD;

PI WPI; 2000-350754/30.

XX N-PSDB; AAD00337.

DR Nucleic acids and encoded proteins involved in the biosynthesis of

XX raffinose, useful for producing soybean seeds with a reduced raffinose

PS content and therefore improved nutritional quality -

XX Claim 22; Page 52-54; 58pp; English.

XX The present sequence is a raffinose synthase from

CC clone wlm24.pk0021.h1 isolated from a wheat seedlings cDNA

CC library wlm24. Raffinose synthase is involved in the biosynthesis

CC of raffinose and higher homologues in the raffinose saccharide family

CC from sucrose. The present sequence is useful for reducing the raffinose

CC saccharide content of soybean seeds which improves the nutritional

CC quality of the soy protein products derived from them.

XX Sequence 751 AA;

Query Match 74.6%; Score 2979.5; DB 21; Length 751;
Best Local Similarity 72.1%; Pred. No. 2.8e-300;
Matches 544; Conservative 94; Mismatches 109; Indels 7; Gaps 5;

OY 1 MTPVPRKISVNDGKLVHGKTLTGVPNDVNIETPGSGRGIVTGAFVGATASHSKSLHVFPM 60
DB 1 mtpvprkisvndgklvngkltlgvpndvnltpgsrgivtgafvgatashskslhvfpm 60
OY 61 GVLGSLRPMCCFRPKLMMTORMGTGRDVPLETOFMLE--SKSEETDGENSPIITYVL 118
DB 61 gvlgsrlpmccfrpkllmmtrmgctgrdvpлетofmle--SKSEETDGENSPIITYVL 118
OY 119 LPLEGQFRAVLQGNDRKNEIETCLESQDNAAVETDQGLHMYVMHAGNPREVINQAVKAVE 178
DB 119 lplegqfrevlqgnrkneietclesqdnaveTdqglhmymhagnprevinqavkave 178
OY 179 KHMOTFLHREKRLPSCLDFGMCWTDAFYTDVTAEGVEEGKSLSQGGTPRFLIIDGMQ 238
DB 179 khmotflhrckrlpscldfgmcwtadafytdvtaegveegkslsqggtpRFLIIDGMQ 238
OY 239 WOQIENKAKDATECLVOEGAOFATRLTGIKENTKFOKLIQNDQNSGLKHLVHGAKQHNHV 298
DB 239 woqiENKAKDATECLVOEGAOFATRLTGIKENTKFOKLIQNDQNSGLKHLVHGAKQHNHV 298
OY 299 VKNYYVNHALAGYWGKVPRAATGMEHVDLAVPYOSPGVLGNOPRIYMDSLAVHGLGLV 358
DB 299 vknYYVNHALAGYWGKVPRAATGMEHVDLAVPYOSPGVLGNOPRIYMDSLAVHGLGLV 358
OY 359 HPKVFENYNEHLAYLASCGVDGKVDVQNIETLGAGHGRVSLRSYHHALEASIASN 418
DB 359 hpkVFENYNEHLaylasCGVDGKVDVQNIETLGAGHGRVSLRSYHHALEASIASN 418
OY 419 FTDNGCIACMCHNTDGLYSAKQAIYASDDFPRDPASHTIHSSVAYNSLLEGFEMQPDW 478
DB 419 ftdngciacmchntdglYsAkQAIYASDDFPRDPASHTIHSSVAYNSLLEGFEMQPDW 478

Db 418 fdngsciscmhntdmlysakqtavrasddfyprdpashvshsavyntclfigefmq 477
QY 479 DDMFSLPADYHAARAIGCCPIYVSDKGNHNPOLLKVLVPGDSVLRADLPGRPT 538
Db 478 dwmfslpadyhaaraigccpiyvsdkpgnhnlkklvlpdgsvlraqlpgrpt 537
QY 539 RSLFVPADRLSLKTLWNKSGVGVFNCOGAGMCKIEKTRHDTSPGTLTASVC 598
Db 538 rdclfsdpardgasllkikvnmkscagvvgvfcngagcvcvkrkrlndeaqcltgsvr 597
QY 599 ASDVDLITQVAAQ-EWLGDTIYVAYRSGEVITLPRKGVISPTLKLFEELHFCPIQIA 657
Db 598 aevveglcqlqddctgdaavvtnhagelivprgatlpvllkleyelthvcpvava 657
QY 658 PSISFAIGLDMFNTGAVEQVEIHNRAATK-TIALSVGRGRGVYSOSRPLKCVYVG 716
Db 658 pdlfsfipglhlmfnagaveecvrtneaddkavalvrvgrgtrgacsrtpakcsids 717
QY 717 AETDFNDESETGLTFISIPVSPDEMTKMSIEIQV 750
Db 718 advefigdadtlgtlvdpvpeemrywlclelrv 751

RESULT 3

AA70975 standard; Protein; 756 AA.

AA70975;

09-AUG-2000 (first entry)

Corn raffinose synthase #2.

Corn: raffinose synthase; raffinose saccharide; soybean; nutritional; soy protein.

Zea mays.

MO200024915-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24923.

23-OCT-1998; 98US-0105451.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Allen SM, Hitz WD;

WPI: 2000-350754/30.

N-PSDB: AAD00332.

Nucleic acids and encoded proteins involved in the biosynthesis of raffinose, useful for producing soybean seeds with a reduced raffinose content and therefore improved nutritional quality -

Claim 2; Page 36-39; 58pp; English.

The present sequence is a raffinose synthase from corn cDNA libraries chn10.pk0034.e8 and chpc24.pk0003.h7 isolated from Raffinose synthase is involved in the biosynthesis of raffinose and higher homologues in the raffinose saccharide family from sucrose. The present sequence is useful for reducing the raffinose saccharide content of soybean seeds which improves the nutritional quality of the soy protein products derived from them.

Sequence 756 AA;

Query Match 71.9%; Score 2874; DB 21; Length 756;
Best Local Similarity 68.8%; Pred. No. 2.8e-289;

Matches 521; Conservative 102; Mismatches 126; Indels 8; Gaps 4;

QY 1 MMYTPKISVNDCKLIVHGGITLITGVDPNVVLPFGSGRLVYGAVGNATASKSLAHFPM 60
Db 1 mlvtpkilsvdgtrilvrgtrvltgpdhvsaaahagagllvgaglvagaagaksnhvltf 60
QY 61 GVLEGLRFKCCFRFLMWTOMTQCGRDVPLETFQFMLESKESKESDENSPIIYTVLLP 120
Db 61 gtlrldrficllfrklwmqtmgtgysgrcdvprletqfmvlvepaadgddap-ayvwm1p 119
QY 121 LLEGGFRALVQGNDRKKEIEICLESQDNVETDQGLHMYHAGTNPFPVINOAKAYEKH 180
Db 120 llegqfralvqgnrdelqiclesgdkavqclqaaahmylhaqndpftvlaaavakekh 179
QY 181 MOTFLREKRLPSCIDMFGMCTMDAFYTDVYAEVEBGLKSLSGGTPRFLITDGMQ 240
Db 180 lqtfhnrdrkklksflfdwfgwctwafyldvdaadvknqslsksggpprflitddgwg 239
QY 241 QIENKAKDAETECLOBGAOFATRLTGIKENTKFOKKLONN----EOMSGLHVLHGAKOH 296
Db 240 qlasenkpdpnvavqeqaiaarltglxentkftqkpdqddgdegaagglkrlyaeekda 299
QY 297 HMYKNTYVNHALAGYWGVP-AATGMEHDTALAYPVOSPGVLGNPDVMDSLAVHGL 355
Db 300 hgvkyvvhmagaywgvtrptaagtamelyepalapyvgpvgvngnqpdlymdslsvlgl 359
QY 356 GLVHPKRVNFNEHAYLASCQVDGVKDVONITETLAGHGRVSLTRSYHHALEASI 415
Db 360 glvhprrvrdfygelhaylasqvgdvkvdqnilletlgaqhgsvraltrayhraleasv 419
QY 416 ASNFQNGCIACGCHNTDGLYSAKOTAIYRASDDYFPDPASHTHISVAYNSLFGEEF 475
Db 420 arsfpngciscmhnschmlysaargtavrasddfyprdpashvshsavyntclfigef 479
QY 476 MOPDMEHSLPADYHAARAIGCCPIYVSDKGNHNPOLLKVLVPGDSVLRADLP 535
Db 480 mqpdcwmfnslhpaaeyhaaraigccpiyvsdkpgnhnlkklvlpdgsvlraqlp 539
QY 536 RPTRDSLFPVDPADRLSLKTLWNKSGVGVFNCOGAGMCKIEKTRHDTSPGTLTASVC 595
Db 540 rptrdslfvpdpardgesllkikvnmkscagvvgvfcngagcvcvkrkrlndeaqcltgs 599
QY 596 SVGASVDLITQVAA-EWLGDTIYVAYRSGEVITLPRKGVISPTLKLFEELHFCPI 653
Db 600 tvradvdalariagdgqgwdgetvyahrtrelrvlprgvalpvlqlgylevhwcp1 659
QY 654 QETAPISFAIGLDMFNTGAVEQVEIHNRAATK-TIALSVGRGRGVYSOSRPLKCV 713
Db 660 ravvpavstapvlldmfnagaveecvldsvgkamalvrgcgrtgaycstrepac1 719
QY 714 VEGAETDFNDESETGLTFISIPVSPDEMTKMSIEIQV 750
Db 720 ldsaevefsdydtglvsvdlrvpegelylwc1elml 756

RESULT 4

AA70977 standard; Protein; 763 AA.

AA70977;

09-AUG-2000 (first entry)

Rice raffinose synthase from clone r1s72.pk0020.d0.

Rice: raffinose synthase; raffinose saccharide; soybean; clone r1s72.pk0020.d0; nutritional; soy protein.

Oryza sativa.

Sequence 763 AA;

Query Match 71.9%; Score 2874; DB 21; Length 756;
Best Local Similarity 68.8%; Pred. No. 2.8e-289;


```

Db      1 mtyssvkvaggselvgtrtvlsgvpeavrrasaagapdvglfgddlaepssrlnvsl 60
QY      61 GVEGLEFMCCFRPKLMMTQRMGTCGRDVPLETOFMLIESKESEDTGENSPITTYVLLP 120
Db      61 gntgmfmecfllkwwagqmgekgyddphtqfllvexxxxx-----xxxxxlvlp 114
QY      121 --LLEGOFRVAVLOGNDK--NEIEICLESQD--NAVETOGLHMYVMAHAGTNPEEYINAA 173
Db      115 pascsegafrrpslsggagadelqlcvesgdagtrraasfdalrtv---gpdsdpfaalaga 172
QY      174 VKAEVHMOTFLHREKKRLPSCLDWFGWCTWDAFYTDVAEGVEGLKSLDSOGTPRRL 233
Db      173 vaaakscilfftraecklpglvdyfgwctwdafygdvtrqegveagllrtstlsggarpkvt 232
QY      234 IIDDGMOOI--ENKADATECLVQEGAOFAFRTLGTIKENTKROKLIQNNQDSGLKHLVH 291
Db      233 llddgwsgvgtldhnpddtgaadkdkpjlartltgikensktg---dgdppaagiktvtvr 289
QY      292 GAKOHNVKNVYVWMAHLAGYWGVRPAATGMEHFDALAYPVQSPGLGNODIYVWDSLA 351
Db      290 aakekyglkyvyywhaitgywgvtrpvvgmgyhsnmgfprvpsgvneepgmktldvlt 349
QY      352 VHGGLVHPRKVFNFYNEHLHAYLASCGVDGVKVDVQNIETLQAGHGRVSLTRSYHNL 411
Db      350 tggltgltvhprrayrfydelhaylaagvdygvkvdyvcclletlgaqgyrvtlrrqfhgal 409
QY      412 EASIAFNFDNGCIACMCNHTDGLYSAKOTAIYRASDDPRPDPAHSHTHISSVAYNSLF 471
Db      410 dasiafnfengliacmshntdalycaakqtaavrasddtprdpvshthlasvaynsyf 469
QY      472 LGEMQPDMDMFHSLHPADYHAARAIGCGPIYVSDKPGNNFDLLKTLVLPDGSVLA 531
Db      470 lgefmjpdmdmfhslhpagdgyhsaraalsgprvysdarpkhnfelkkmvlpdgsvlra 529
QY      532 QLPGRTRBSLVDPARDRTSLIKTNLKNKCGVGVFNCGAGWCKIEKTRIHDTSPG 591
Db      530 wlpgrtrbslvdpardrtsltktnlknkcgvgvfnccgagwckiekttrihdtspg 589
QY      592 TLTAASVCAADVDLITQVA--GAEMLGDTIYAYRSGEVLRLPKVGSIPVTLKYLEFELFH 650
Db      590 alscgykgsdvhliadaatdsemdgcavyrnasadlvlpnqaalpslklvlehdlltv 649
QY      651 CPLOETAPSSISPAIGCLDMENTGAVEQVEIHNRAAKT----- 690
Db      650 spikdlapgftrfaplglyvdmfinsgaavegltyhrlidykvsangsastlpeqlsagql 709
QY      691 --IALSVGRGRFEGYSSORPLKCVVGGAEFNDSEGLTTSIPVSPPEE 740
Db      710 gltvcmevrgcgkfgayssvtrpckmlgsaqveftydsassgvlvldlempke 761

```

```

RESULT 6
AAV70981
ID AAV70981 standard; Protein; 841 AA.
XX
AC AAV70981;
XX
DT 09-AUG-2000 (first entry)
DE
DE Wheat raffinose synthase from clone wlm96.pk033.h5.
XX
KM Wheat; raffinose synthase; raffinose saccharide; soybean;
KM clone wlm96.pk033.h5; nutritional; soy protein.
XX
OS Trilicium aestivum.
XX
PN WO200024915-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24923.
XX

```

```

PR      23-OCT-1998; 98US-0105451.
PA      (DUPO ) DU PONT DE NEMOURS & CO E.I.
XX
PI      Allen SM, Hiltz WD;
XX
DR      WPI: 2000-350754/30.
DR      N-PSDB; AAD00338.
XX
XX
XX      Nucleic acids and encoded proteins involved in the biosynthesis of
PR      raffinose, useful for producing soybean seeds with a reduced raffinose
PR      content and therefore improved nutritional quality -
PS      Claim 2; Page 55-58; 58pp; English.
XX
XX
XX      The present sequence is a raffinose synthase from
CC      clone wlm96.pk033.h5 isolated from a wheat seedlings cDNA
CC      library wlm96. Raffinose synthase is involved in the biosynthesis
CC      of raffinose and higher homologues in the raffinose saccharide family
CC      from sucrose. The present sequence is useful for reducing the raffinose
CC      saccharide content of soybean seeds which improves the nutritional
CC      quality of the soy protein products derived from them.
XX
SQ      Sequence 841 AA;

```

```

Query Match 52.88; Score 2110.5; DB 21; Length 841;
Best Local Similarity 53.18; Pred. No. 1.1e-209;
Matches 407; Conservative 123; Mismatches 213; Indels 23; Gaps 10;

```

```

QY      1 MTVPRKISVNDGKLVNHRKTLTGVPDNNVLLTPSGGGLTVGARGVATASHSKSLHVFPM 60
Db      81 mtiesvvljaggeislvgtrtvlsgvpaavsaaspaarqpvdyglgdlagparthvsl 140
QY      61 GVEGLEFMCCFRPKLMMTQRMGTCGRDVPLETOFMLIESKESEDTGENSPITTYVLLP 120
Db      141 ghmvgtrfmacfllkwwagqmgekgyddphtqfllve---alsgeeddasvyyvflp 198
QY      121 LLEGOFRVAVLOGNDK--NEIEICLESQD--NAVETOGLHMYVMAHAGTNPEEYINAA 175
Db      199 lvegafraslsggagadelqlcvesgdagtrlaasfdalrtv---gaadsdpfaalagava 256
QY      176 AVEKHMOTFLHREKKRLPSCLDWFGWCTWDAFYTDVAEGVEGLKSLDSOGTPRRLII 235
Db      257 avzscldgtftrpaecklpalvdyfgwctwdafygdvtrqegveagllrtstlsggarpkvt 316
QY      236 DDGMOOI--ENKADATECLVQEGAOFAFRTLGTIKENTKROKLIQNNQDSGLKHLVHGA 293
Db      317 ddgwsgvgtldhnpddtgaadkdkpjlartltgikensktgsg-dpdatatgletlvtara 375
QY      294 KOHNHNVKNVYVWMAHLAGYWGVRPAATGMEHFDALAYPVQSPGLGNODIYVWDSLA 353
Db      376 kekylkyvyywhaitgywgvtrpvvgmgyhsnmgfprvpsgvneepgmktldvltlq 435
QY      354 GLGLVHPRKVFNFYNEHLHAYLASCGVDGVKVDVQNIETLQAGHGRVSLTRSYHNLAE 413
Db      436 glglvhpqavhrfydelhaylaagvdygvkvdyvcclletlgaqgyrvtlrrqfhgalda 495
QY      414 SIASNFNDNGCIACMCNHTDGLYSAKOTAIYRASDDPRPDPAHSHTHISSVAYNSLFTG 473
Db      496 svaknfdngliacmshntdalycskqtlavrasddtffceavshthlaaavaynsfvlj 555
QY      474 EFMOQPDMDMFHSLHPADYHAARAIGCGPIYVSDKPGNNFDLLKTLVLPDGSVLA 533
Db      556 efmlpdcvdmfhslnhpagdgyhsaraalsgprvysdarpkhdelftkmvlpddtvtlratl 615
QY      534 PGRPTRBSLVDPARDRTSLIKTNLKNKCGVGVFNCGAGWCKIEKTRIH-DTSPGT 592
Db      616 pgrptrbslvdpardrtsltktnlknkcgvgvfnccgagwckiekttrihdtspgt 675
QY      593 LTASVCAADVDLITQVA--GAEMLGDTIYAYRSGEVLRLPKVGSIPVTLKYLEFELFH 649
Db      676 ltcvtrgrdvhliaaatdggagsgdcavyrhngagdlvlpdgvalpslklvlehdvlt 735

```


Best Local Similarity 39.7%; Pred. No. 5,6e-139;
Matches 302; Conservative 127; Mismatches 262; Indels 69; Gaps 15;

```

QY 8 SYNDKLVHGHKTLITLGPVNVLTLPSSGRLVTGAFVGTASHSKSLHVPNGVLEGLR 67
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 22 slcnstlkvngqvllsqvknvlltpctyhtgtcflgtfatspkrharpjqlknis 81
QY 68 FMCFERKLMWMTQRMGTGRDVPLETOFMLIESKESETDGENSPILYVTLTLBLEQGR 127
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 82 ftsifkrkvwtltwtgsnrdteteqfmlimgshp-----yvfllrllqpptr 130
QY 128 AVILOGNDKNEIEICLESQDNVAVETDGLHVMYHAGTNPREVINOAVKAVEKHMOTFLHR 187
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 131 aslqpheddvavcevsghvlassfdvtvylhagdnprtlvkeamrvrvalgsfkl1 190
QY 188 EKKRLPSCLDWFGMCTMDAFTYDTVTAEGVEGLKSLSGQTPPRFLIDDGMOQIENKAK 247
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 191 eekcvpwmvdkfjwctwclafyltvhpqyregvklvdggcpqpfyllddgwqclshdsd 250
QY 248 DATECLVQ--EGAQFATRLTGIKENTFKOKLONNEQMSGLKTLHVGAKOHN--NVKNVYV 304
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 251 pekegmmtqvagegmprcllsyeenykf-----tsykegklkfgvrelkeefsgvevyv 306
QY 305 WHALAGVGVKPAATGMEHYDTALAVPVQSPVLGNQPDIVMSLAVHGLVHPKRYV 364
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 307 whalcygwgyvtrpvagvma--eaavekprklteglkgmedlavdklvnngvgyvvpelvy 364
QY 365 NFNELHAYLASGVGVKVDYQNIETLGAAGGRVSLTRSHALHLSIASNFMDNGC 424
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 365 emyegjlahlesagldgvkvdvllhlemvcekygvgvdmakaykaltasvrvhfkngnyv 424
QY 425 IACMCHNTD--GLYSAKOTATVIRASDDEYPRDP-----ASHTHISSVAVNSLFL 472
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 425 laamehondfmltltgaetalsigrvcdftwcdpdygdpngtflwsgclmnhv--caynsllwm 481
QY 473 GERMOPDMWFHSLHRAADYHAAARIGCPTIVSCKPGNHNFDLKLKLVLPDGSVLAQA 532
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 482 gnlihpdmfngtclpccaahaasraalsgprlylsdvtgnhfellktlalpdsallrce 541
QY 533 LPRPTRDSLFVNPARDRTSLKTLWLNKSGVGVGFVNGOGAKWCKIEKTRHDSPT 592
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 542 hyalprtdcfdadplndgktmlkikwlnkytgyivfncqggwtr-----e 588
QY 593 LNASVCASD---VDLITVAGAEW-----LGDITVAVRSGE---VIRLPKGVSI 636
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 589 lrsnkcaaeishrvstkltkldiewdsgknplsliegvtlfaayfsgakklillsapsdde 648
QY 637 PVTLKYLELELHFHFCIOEL-APSISFAALGLIDMNTGCAVEQVEIHNRRAAKTTLASV 695
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 649 ealslepfntellitvsvptvlpqksvktapldglvmlntlgavqslafde--gqnllvevjl 706
QY 696 RGRGRGVYSSQRLPKCVGAGVGAETDENYDSEGLTTFESIP 735
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 707 rgtgemrvyasekptrctldgkevdlcy--egsmwnlqyvr 744

```

RESULT 9
AA32074
ID AA32074 standard; Protein: 777 AA.
XX
AC AA32074;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mustard raffinose synthase.
XX
KM Raffinose synthase; mustard; transgenic plant.
XX
OS Brassica juncea.
XX
FH Key Location/Qualifiers
FT Misc-difference 210 /note="encoded by ACR"
FT

```

XX XX EP953643-A2.
PN 03-NOV-1999.
PD 27-APR-1999; 99EP-0107430.
PF 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX Watanabe E, Oeda K;
PI WPI, 1999-593144/51.
DR N-PSDB; AA220209.
XX
PT New sense and antisense genes, useful for altering the level of
PS raffinose in food plants -
XX
PS Claim 26; Page 29-31; 55pp; English.

```

This sequence represents mustard raffinose synthase, a protein that can bind a D-galactosyl group through an alpha(1-6) bond to the hydroxy group attached to the carbon atom at the 6-position of the D-glucose residue in a sucrose molecule to form raffinose. cDNA (see AA220209) encoding the enzyme was isolated from mustard (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from plant raffinose synthase genes (see AA220207-10) may be used to obtain other raffinose synthase genes by labeled detection or amplification (claimed). These genes may be used to control the levels of raffinose produced in plants. Antisense genes can be used to knock out existing gene activity, and sense genes to increase the level of gene activity. The resulting transgenic plants may be used as a food source to alter the growing conditions for gut enterobacteria, providing general health advantages.

Sequence 777 AA;

Query Match 35.4%; Score 1415.5; DB 20; Length 777;
Best Local Similarity 38.2%; Pred. No. 2.1e-137;
Matches 296; Conservative 133; Mismatches 279; Indels 67; Gaps 17;

```

QY 5 KRISVNDKLVVHGKTLITLGPVNVLT-----PSSGRGLVTGAFVGTAT-SHKSLS 55
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 22 plrlfegsdllanghvvltvprvntvlaspyladkgepydasagsfifnldgeprsr 81
QY 56 HVPNGVLEGLRPFMCCEFRKLMWMTQRMGTGRDVPLETOFMLIESKESETDGENSPILY 115
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 82 hvastlgklrldrtfmsliffkvwctlwvgsksdsdienetqlillen-----ssggrp--y 134
QY 116 TVLLPLEEGOPRAVLQSGNDKNEIEICLESQDNVAVETDGLHVMYHAGTNPREVINOAVK 175
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 135 vlllprllfegsfrsftqgedddavavcesgsvtgytsefgyvvyvnhagddpklvkdamk 194
QY 176 AVEKNMOTFLHREKKRLPSCLDWFGMCTMDAFTYDTVTAEGVEGLKSLSGQTPPRFLII 235
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 195 vrvvhnmtfklleekltppglvdkfgwctwdaftltnpddvghkvkclvdggcpqpfvl1 254
QY 236 DDGMOQIENKAK----DATECLVQEGNQFATRLTGIKENTFKOKLONNEQMS-GLNHLY 290
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 255 ddgwsqslhdsdgidvegmsctv-ageqmporllkfgentkftdyvprkdkneymkatfv 313
QY 291 HGAKOHHN-VKNVYVWHALAGVGVKPAATGMEHYDTALAVPVQSPVLGNQPDIVMS 349
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 314 rdlkeefstvdvlyvwhalagvgygqlrppgplrp--pstlvrepelspglltmqdlavdk 371
QY 350 LAVHGLGLVHPKRYVFNELHAYLASGVGVKVDYQNIETLGAAGGRVSLTRSHYH 409
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 372 lvdgtgltfsvpmanefyegjshnlgnvgyldgvkvdvllhlemvcekygvgvdmakayfk 431

```

OY	410	ALEXISNENFNDNCIACMCNNDGVL-SAKORAIYASDFPDRDAS-----	457
		: : : : : :	
Db	432	altssvkhfngvismetmelndfmllgicailsgrvgddfwctdpsglnftyLgqc	491
OY	458	HTHISSVAVNSLFLGEPMQPDMDHSLHPADYHAAAPAGGCIYVSDKRGNNFDL	517
		: : : : : :	
Db	492	hmvh--caynsllmngnflfgpdwmtfstbpcaeaaasaisgpyliysdcvgghdfl	548
OY	518	LKRLVLPDGSVLRQLQCRPRDLSLPDAPARDRLSLKINMLNKGSGVGVFVFCQAGWC	577
		: : : : : :	
Db	549	lkrlvlpdgsllrrehyalprtdrlfeadplhdygtlmklvmlktygilafgncggagwc	608
OY	578	KIEKTRIHDTSPGTLRAASVADLTVOYAAEW-----LGTIYAYRSGE	626
		: : : : : : : :	
Db	609	retrrngqsfqcvlllatatlnpkdv-----ewmsgnplsvenveefalflsgsk	659
OY	637	VIRLPKGVSIPTVLKLEFEELHHCPIQIEI-APSISFAAIGLDMFTMGAVAEVETLHNR	685
		:: : : : : : : : : : : : : : : : : : :	
Db	660	lvlsqpddelettlepfkfeiltvsvpytlsegssvqfapqlylmmltsgafrslryhe-	718
OY	666	AATKTIALSVGRGRFGFYSSQRPLKCVGGAELDFWYDSTGTLTTSIYVSPEE	740
		:: : : : : : : : : : : : :	
Db	719	--esveelvgrvgaefvayarskascldgevefey--eesmvmvypmsape	768

RESULT	10
AAW53570	
ID	AAW53570 standard; Protein; 784 AA

DT	06-JUL-1998	(first entry)
XX		
XX	Cucumber raffinose synthase.	
XX		
XX	Cucumber; raffinose synthase; sucrose; galactinol.	
XX		
OS	Cucumis sativus.	
XX		
PN	JP10084973-A.	
XX		
PD	07-APR-1998.	
XX		
PF	28-APR-1997; 97JP-0111124.	
XX		
XX	26-JUL-1996; 96JP-0198079.	
PR	26-APR-1996; 96JP-0107682.	
XX		
PA	(AJIN) AJINOMOTO KK.	
XX		
XX	WPI; 1998-264858/24.	
DR	N-PSDB; AAV22250.	
XX		
PT	Raffinose synthase gene - useful for preparation of raffinose in transformed plant	
XX		
PS	Claim 3; Pages 17-20; 26pp; Japanese.	
XX		
CC	The present sequence is cucumber raffinose synthase, which	
CC	forms raffinose from sucrose and galactinol, has an optimum pH of	
CC	6 to 8 and working temperature of 35 to 40 degrees C, has a	
CC	molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa	
CC	by PAGE and SDS-PAGE under reductive conditions and is inhibited by	
CC	iodoacetamide, N-ethylmaleimide and myoinositol.	
XX		
SQ	Sequence 784 AA;	

Query Match	34.98;	Score 1394;	DB 19;	Length 784;
Best Local Similarity	37.68;	Pred. No. 3.7e-135;		
Matches 289;	Conservative 143;	Mismatches 285;	Indels 52;	Gaps 15;

```

0Y 8 SVN0GLVNHCKITLLEQVPRNNVLTRESGGL-----YGFANFQASASKSLHNEPMV 62
Db 26 aiddsgdfvnghsflsdvpenlvaaspryslksprvsyvcfvgffaaesdrfthvsyjk 85
0Y 63 LEGRFMCCEFRFKLMMWTORMGTCGRVPLETOFMLESKESETDGENSPITVTLPL 122
Db 86 lkdifmslffkvwetlhwgyngrglsestqvlleked-----gpr--yvfllprlv 138
0Y 123 EGQFRAVLQNDKNEIELESQDNAVEPDQGLHMYTMACTPREVYNQAKAVEKNQ 182
Db 139 egfrftrsqdddfvdvcevsqskvvdasfrsmlylhagddpalfvkaamkivrlthlg 198
0Y 183 TFLHREKRRLPSCLDMWGMCTWDAFEYTDYAAEGBEELKLSQGGPRPRLLIDDDMOOI 242
Db 199 ftllleektrpgrjvdklqvcwdaifyltvprgyalevznlvddgcerpavliddgwgsl 258
0Y 243 ENKADATECJVOE---GAOFATRLTQIKENTKFO-----KRLONNOMSGLNHLNGAK 294
Db 259 ghdsprlckegmqtvgsegmrcrlllkfgenykrtdyvnprkatgrsgdqgymkafidelk 318
0Y 295 -QHNNVKNVYVWHAALAGVSGVKPRATGEMEHYDTALAYPOSGEVLGNORDYIMDSLAVH 353
Db 319 gefkvevhuywhalocgywglrtpqgrlp--earvlyqrlspglqmtledalvdvklvhl 376
0Y 354 GIGLVHPKKVNFNFNEHLNATYASGCVQGVKVDQVONIIETLGAGNGGVSLTRSYHNALEA 413
Db 377 kyglvprpkaeemqgylahlekvgldrvkldvhlhlemlecedygrfvdlakayukamtk 436
0Y 414 STASWPTONGCIACGCHNTDCLY-SAKOTAIVARASDDEYPRDPAS-----HTIH 461
Db 437 slnkhflkngvylaamehcnndmfmgltgeaislgyrgddftwctdpsqgrngftwlgscimwh 496
0Y 462 ISSVAYNSLFTGEFMQPDMDFNHSLHPRADYHAARAIGCCPIYVSCKPCNNHFNDLKL 521
Db 497 ----candslmgngflhpqwdmfgsthrpcaafhaasraaisgprlyvdsavgkhmfdllkl 553
0Y 522 VLPDGSVLRADLPGRFPTDLSFLVPRPARNRSTLSLITWMLNKGSGVGVGVEGOCGAWCKIEK 581
Db 554 vlppdgslltrseyalprtdclfedrplhngelmklwmlnkftgyigaifncqggwccetr 613
0Y 582 KIRINDTSPGLTASVCASVDVLT--OVAGAEMVLGPTIYAVAYSGSVIRLPKGVSPVPT 639
Db 614 mqcfsqykrstxtpkcdlewmhsgenpislgyvktfallyakaklllstkpsqddia 673
0Y 640 LKVLFEFLHFNCPIOE-IAPISFAIGLDMFNTGCAVEQVETHNRAAKITLALSVGR 698
Db 674 ldrfefelltvsprklqltstlhfaripjlymnltsaigaigvdvddds--sveigvkc 731
0Y 699 GRFGVYSSGRPLKCYVVGAEIDFNDVDSGTGLTFSIPSPSEMYRWSLE 747
Db 732 gsmryfaskkpractldegedvfykvdqdg-muvvqvr-----wpid 771

```

RESULT	11
AAV17417	
ID	AAV17417 standard; Protein; 784 AA

DT	29-JUL-1999	(first entry)
XX		
DE	Cucumber	raffinose synthase.
XX		
KW	Raffinose synthase;	sucrose; galactinol
XX		
OS	Cucumis	sativus.
XX		
PN	JP11123080-A.	
XX		
PD	11-MAY-1999.	
XX		
PF	24-OCT-1997;	97JP-0292969.
XX		


```

PR 24-OCT-1997; 97JP-0292969.
XX
XX (AJIN ). AJINOMOTO KK.
XX
XX WPI: 1999-340516/29.
DR N-PSDB; AAX61238.
XX
XX New raffinose synthase gene - for production of raffinose from
PT sucrose and galactinol
XX
XX Claim 2; Page 25-27; 37pp; Japanese.
XX
XX The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.
XX
XX Sequence 784 AA;
SQ

Query Match 34.9%; Score 1394; DB 20; Length 784;
Best Local Similarity 37.6%; Pred. No. 3,7e-135;
Matches 289; Conservative 143; Mismatches 285; Indels 52; Gaps 15;

OY 8 SVNDGKLIVHGKTLITGVDPNVYLPFGSGRGL-----VTGAFVGATASHSKSLHPEPMGV 62
   :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 26 aIdgsdftvngsfldsvpenlvaasprytsIdksrsvsgctvgfdaSepdsrthvvsIqk 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 63 LSGLRPMCCFRKLMWMQRMGTCGRDVPLETOFMLIESKESSETDGENSPILYIVLPL 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 86 lkdrtmslfrkvwthwgrngdlesetqlvleKds-----grp--yvflrlplv 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 123 EGOFRANVLOGNKNEITELCSGDNAVETDGLHWYVHAGNPREVINOAKAVEKHMQ 182
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 egrftrtsIdpgddfdvdvcesgsKvvdasfrsmlylhagddpfaIvkeamkIvrtlyg 198
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 183 TELHREKKRLPSCLDWFGCTWDAFYTDVTAEGVEGLKSLSGSGTPPRFLIDGMOOI 242
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 tfrllleekrpprlvdkfvgctwdaIyIvhpqvgIegvgrIhIvdggcpgpIvIddgwgsI 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 243 EKKADATPECLVQF---GAQFATRLTGIRKENTKFO-----KKIQNNEQMSGKLHLVHAK 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 259 gIdsprrltkegmngtvaageqmpcrlIkIqenykfrdygnrkatgragqkmaKfaIdelk 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 295 -OHNNKVVYVHAGVAGVYKRAATGMEHDTALAPVQSPGVLGNQRPVMDSLAVH 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 319 gelftevhyyvwhalcgywgglrtpvprpIv--eaivIqrvIspglqtmMedIavdkIvIh 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 354 GLGLVHPKRVFNFNELHAYVLASCGVDGVKVDVONIETLGAGHGGRVSLTRSYHHALEA 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 377 kvglvprpekaemyegIhahlekvgIdgYkIdvIhIhIemlecedyggrrIdakaykkanKt 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 414 STASFTDNGCJACMCNHTDGLY-SAKQTAIVRASDDFYPRDPAS-----HTIH 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 437 slnkIfkIqngvIasmeHndfmlqItealsIgrvgddfwctdpsgdprgtfwIggchmvh 496
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 462 ISSVAVNSFLCEEFQOPMDMHSLHPADYHAARAAIGCGPIYVSDKRGHNHNDLKL 521
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 497 ---candslwmgntfIhnpwdmfqschpcaahaastraIs9pIvIvdsdvgkInfdIIL 553
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 552 VLPDGSVLRADLPGRPRDLSLFVDPARDRTSLKIMNLNKSQVGVVGNCGAGCKITEK 581
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 554 vIpdgsIrlseyalprtdclfeDplhngemIkIwlnkItgylgaIhncggvgwccetr 613
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 562 KTRINDTSPGILTASVCSDVLI--QVAAEWLIGDTIYVAYRSGEYIRLPKGVSIPTV 639
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 614 tngcfsgyKsvtstKnpdkIewhsgenrlsIegvktfaIyIyqakklIskpsqddIa 673
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 640 LKVLFEELFHPCIOE-LAPSTISFAIQLDMFNTNGAVEVEIHNNRAATITLASRGR 698
   | ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
DB 674 ldrfefeIItvsprvklIqtsIhfarIglvnmIIntsgaIsgvdyddIs--sveIgyvKgc 731
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY 699 GRFGYSSQRPRLKCYVGAETDENTDSTGLTTSIPSPPEAMRWISIE 747
   | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 732 gemvIaaskpracrIdgedvgfkydgdq-mvvvgvp-----wpId 771
   : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
AAM57887 standard; Protein; 781 AA.
ID AAM57887
XX
XX AAM57887;
XX
XX 23-SEP-1998 (first entry)
XX
XX Soybean raffinose synthetase.
DE
XX Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; soybean.
KW
XX Glycine max.
XX
XX EP849359-A2.
XX
XX 24-JUN-1998.
XX
XX 18-DEC-1997; 97EP-0122417.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Oeda K, Wantanabe E;
XX
XX WPI: 1998-324670/29.
XX
XX N-PSDB; AAV40801.
XX
XX New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
XX Claim 1; Page 31-34; 44pp; English.
XX
XX This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
XX Sequence 781 AA;
SQ

Query Match 34.7%; Score 1385; DB 19; Length 781;
Best Local Similarity 38.5%; Pred. No. 3.2e-134;
Matches 293; Conservative 129; Mismatches 273; Indels 66; Gaps 16;

OY 7 ISVNDGKLIVHGKTLITGVDPNVYLPFG-----SGRGL-----VTGAFVGATASHSKSLHAY 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 25 ItlgsnflangnpfltevenlIvprsdaksKnneddvdvgctfvgfhadeprsrh 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 58 FPMKVLGELFPMCCFRKLMWMQRMGTCGRDVPLETOFMLIESKESSETDGENSPILYIV 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 85 asIqkIrgIkfmslIfIkvwthwvsgnhelhehtqmmIId---kndqIgrp--fvI 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 118 LRLPLEGQFRAVLVQGDNDKNEITELCSGDNAVETDGLHWYVHAGNPREVINOAKVAV 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 lIprIlgasIrasIqprIdIvvdvcmesgstrvcgsfscIyvvhgndpqlIreatkvv 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 178 EKIMQFTLHREKKRLPSCLDWFGCTWDAFYTDVTAEGVEGLKSLSGSGTPPRFLIID 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 08:55:37 ; Search time 14.71 Seconds
(without alignments)
1245.357 Million cell updates/sec

Title: US-09-810-186-1

Perfect score: 3996
Sequence: 1 MVTLPKISVNDGKLVHGKT.....TFISPYSPREMYRMSIEIQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2-6/ptodata/2/1aa/5A.COMB.pep.*
3: /cgn2-6/ptodata/2/1aa/5B.COMB.pep.*
4: /cgn2-6/ptodata/2/1aa/6A.COMB.pep.*
5: /cgn2-6/ptodata/2/1aa/6B.COMB.pep.*
6: /cgn2-6/ptodata/2/1aa/backfile1.pep.*

. Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1394	34.9	784	US-08-846-234-5	Sequence 5, Appl1
2	119.5	3.0	1074	US-09-004-838-111	Sequence 111, App
3	117.5	2.9	4544	US-08-469-486-52	Sequence 52, Appl
4	117.5	2.9	4544	US-08-469-658-52	Sequence 52, Appl
5	115	2.9	478	US-09-004-838-51	Sequence 51, Appl
6	109.5	2.7	780	US-08-232-538-14	Sequence 14, Appl
7	109.5	2.7	780	US-08-786-164-14	Sequence 14, Appl
8	109.5	2.7	1338	US-08-750-141A-3	Sequence 3, Appl1
9	109	2.7	477	US-09-004-838-46	Sequence 46, Appl
10	109	2.7	477	US-09-004-838-99	Sequence 99, Appl
11	106	2.7	30	US-08-846-234-1	Sequence 1, Appl1
12	103.5	2.6	587	US-08-931-608A-4	Sequence 4, Appl1
13	102	2.6	1817	US-09-004-838-125	Sequence 125, App
14	102	2.6	2206	US-07-852-260-2	Sequence 2, Appl1
15	102	2.6	2206	US-08-461-503-2	Sequence 2, Appl1
16	100	2.6	2206	US-08-465-250-2	Sequence 1, Appl1
17	100	2.5	758	US-08-874-678-1	Sequence 1, Appl1
18	100	2.5	758	US-08-643-839-1	Sequence 1, Appl1
19	100	2.5	758	US-09-051-363-24	Sequence 24, Appl
20	99.5	2.5	566	US-08-484-993B-41	Sequence 41, Appl
21	99.5	2.5	566	US-08-484-158B-41	Sequence 41, Appl
22	99.5	2.5	566	US-08-484-596A-41	Sequence 41, Appl
23	99.5	2.5	566	US-08-480-150A-41	Sequence 41, Appl
24	99.5	2.5	566	US-08-458-731-41	Sequence 41, Appl
25	99.5	2.5	566	US-08-149-223A-41	Sequence 41, Appl
26	98.5	2.5	566	US-08-340-011-5	Sequence 5, Appl1
27	98.5	2.5	1311	US-08-901-710-5	Sequence 5, Appl1

28	98	2.5	750	6	5457037-3	Patent No. 5457037
29	98	2.5	1477	1	US-08-038-682-4	Sequence 4, Appl1
30	98	2.5	1477	1	US-08-302-832-4	Sequence 4, Appl1
31	98	2.5	1477	2	US-08-530-198-4	Sequence 4, Appl1
32	98	2.5	1477	2	US-08-469-880-4	Sequence 4, Appl1
33	98	2.5	1477	2	US-08-728-470-4	Sequence 4, Appl1
34	98	2.5	1477	2	US-08-617-697-4	Sequence 4, Appl1
35	98	2.5	1477	2	US-08-719-641-4	Sequence 4, Appl1
36	96.5	2.4	4303	2	US-08-460-751-2	Sequence 2, Appl1
37	96	2.4	776	4	US-09-346-237-4	Sequence 4, Appl1
38	94.5	2.4	610	2	US-08-974-565C-9	Sequence 9, Appl1
39	94.5	2.4	610	3	US-09-255-748-9	Sequence 9, Appl1
40	94	2.4	751	6	5457037-5	Patent No. 5457037
41	94	2.4	776	4	US-09-346-237-7	Sequence 7, Appl1
42	94	2.4	1004	4	US-09-268-347-30	Sequence 30, Appl1
43	94	2.4	4302	3	US-08-658-136-5	Sequence 9, Appl1
44	93.5	2.3	610	2	US-08-942-521B-7	Sequence 7, Appl1
45	93	2.3	1198	4	US-09-199-637A-405	Sequence 405, App

ALIGNMENTS

RESULT 1
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5

Query Match 34.9%; Score 1394; DB 4; Length 784;
Best Local Similarity 37.6%; Pred. No. 5, 9e-146;
Matches 289; Conservative 143; Mismatches 285; Indels 52; Gaps 15;

QY 8 SVNDGKLVHGKLTILGVPDNLVLTGSGRGL-----VTGAPVGAATASKSLHVPKGV 62
DB 26 AIDGSDFTVGHGFLDLPDNIASPSPTYSIDKSPVSGCVGFGFASPSDRHVSICK 85
QY 63 LEGIRFWCCFRFLMMWMTQMGTCGRDVPLETFQFMLESKESEFTDGENSPIITYVLLPL 122

```

Dh      86 LKDIRFSIFSRKFWMTTHWVGRNGDLESEOTJVLERSDS-----GRB--VYFLPIV 138
      123 EGOIRAVLOGNDKNEIEITCLESNDNAVETOOGLHMYTMAAGTNPFEKYNQAVAVENKNO 182
      139 EGPERTSIQPDDEFFVJVCVESSSKVYDASFRSMYLAHGDDPFAVYKAMKTVRHLIG 198
      183 TFLREKRLRPLSCDMEFGWCTWIDAFYDYVAEGVEEGIKLSOGGUPRFLIIDWMOOI 242
      199 TFRLEKTRPGYIDKRGKWCWPAFYLVTRPGVIGEBVRRLYDGGCPRGVLVIDDGMOSI 258
      243 ENKAKDAITECLVDE---GAOPATRLGICIKENTKFO-----KLIQNNQMSGLNHLVGA 294
      259 GHSDPPTIKESMNQTVAGEQWPCRLTFOENYKFRDYVNRKATGPBAGOKMAFIDELK 318
      295 -QHNNVKNVYVWHLAGWGWVKAATGMEHYDALYVPOSGVLGNQODIYWDLSAVH 353
      319 GEFTVEHYVWHLACGWGSLRGVYGLP--EAKYQPLVSLPQLOMTMEDLADKTVLH 376
      354 GLGLVHRKVFNFENYLNHAYLASCGVDYKVDVONIETLGAHGGRVSLTRSYHALEA 413
      377 KVGVLPEPEKAEEMEGYELHAEKYGIDGVKIDVYIHLLEMCEJDGGHVDLAKAYKAMTK 436
      414 SIANFNDNGICIAOMCNTDLSY--SAQOFAVARSDDFYRDPAS-----HTIH 461
      437 SINHFPGNGVYLAEMECNDMEFGTATISIGRVDDFWCTDPSGDPNGTFWLOGCMVH 496
      462 ISSAVYSLFEGEEMQOPDMDFHSLHBAADYHAARAIGCCPIYSDSPKPNHNFLLKTL 521
      497 ---CANSLMNGNIIHPDMDFQSTHRCAPFAHSAIRASIGPIYSDSVGKHNFDLLKTL 553
      522 VLPQGSVLRALOPRRPRDSITFVDPANDRISLKTIMWLKCSGVYGVFNCOGAGWKIEK 561
      554 VLPQGSILRSEYVLPTRDCLFEDPLHNGEFTMLTINWLNKFTGYIGAFNCOGGWCHETR 613
      582 KTRIHDSPGGLTASVCAVDVLLT--QVAGAEMVLAGDTIYAVYSGAVIRPKGVSPVT 639
      614 RNQCFQSQRKRVKTKNPKDIEHMSGENPISIEGVKTFALTYOAKTLILSKBQDDIDIA 673
      640 LKVLFELEFHEHCPIQE-IAPSISFAAIGLLDMFNTGAVBOVELHNRAATKTIALSVGR 698
      674 LDPEEFELITVSPVTKLIQTSILHAPJIGVLMNLTSGAISVDYDDLDS--SVEIGVKG 731
      699 GRFCVYSSQORPLKCVGGAETDFNYDETGTLTTSIVSPPEKMRWSIE 747
      732 GEMRFASKRPRACRIDEGYFKYDDQD--WVVVVP-----WPID 771

      RESULT      2
      US-09-004-838-111
      Sequence 111, Application US/09004838
      Patent No. 6350933
      GENERAL INFORMATION:
      APPLICANT: Michelmere, Richard W.
      APPLICANT: Meyers, Blake
      TITLE OF INVENTION: Procedures and Materials for
      TITLE OF INVENTION: Confering Pest Resistance in Plants
      NUMBER OF SEQUENCES: 140
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: Townsend and Townsend and Crew LLP
      STREET: Two Embarcadero Center, Eighth Floor
      CITY: San Francisco
      STATE: California
      COUNTRY: USA
      ZIP: 94111-3634
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/004,838
  
```

```

1      FILING DATE: 09-JAN-1998
2      CLASSIFICATION: 800
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: US 08/781,734
5      FILING DATE: 10-JAN-1997
6      ATTORNEY/AGENT INFORMATION:
7      NAME: Einhorn, Gregory P.
8      REGISTRATION NUMBER: 38,440
9      REFERENCE/DOCKET NUMBER: 023070-078810US
10     TELECOMMUNICATION INFORMATION:
11     TELEPHONE: (415) 576-0200
12     TELEFAX: (415) 576-0300
13     INFORMATION FOR SEQ ID NO: 111:
14     SEQUENCE CHARACTERISTICS:
15     LENGTH: 1074 amino acids
16     TYPE: amino acid
17     STRANDEDNESS:
18     TOPOLOGY: linear
19     MOLECULE TYPE: protein
20     FEATURE:
21     NAME/KEY: -
22     LOCATION: 1..1074
23     OTHER INFORMATION: /note="RG2K deduced sequence"
24
25     US-09-004-838-111

```

Query Match	3.0%	Score 119.5	DB 4	Length 1074
Best local Similarity	18.9%	Pred. No. 0.0023		
Matches 130	Conservative 94	Mismatches 261	Indels 203	Gaps 29
QY 133	NKNEIEICLESADNAVETDQGLHANYMHA	GTNDFEYINQAYAVAKEMHQTFLH-----	186	
DB 150	NDFESREHFFRKALFELAGISNHTSHVVALWGMGVGKTTMMRLNIIKERTHRYIYLVV	-----	209	
QY 187	REKRLRPSCLDFMFCWTDFAEYTV-----	TAGVEEGKLSKSOGATPREPFLIID	236	
DB 210	IKEMMDLISIQDAV-----	ADYLDKLTESNESTRAKLKEGGQAQSDGKRNFLIID	263	
QY 237	DGMQOIE-----	NKADATECLVQEGAFATRLTGIKENTFKQKKLQNNQMSGL	286	
DB 264	DVMQSVMEDIGISPPRNOGVGFKYLITSNKDYCAKM-GVEANLIPDYVFLTEEEHOSLT	-----	322	
QY 287	KHLVHGAKOHNNKANYMHALAGVAGV-----	KPAAGMEHED-----	326	
DB 323	FYQGFVKVSTHDK-----	IGKATVRNCGGPIAKTIANTLKRNKKDYKQKSLSIENHDI	379	
QY 327	TALAPV-----	QSG-PGYLGNQDPIDVMSLAVHGLGVHPRKVFN-----	365	
DB 380	ETIAHVQMSYDNLQNEAQSIFLLCGFPRFEDFIPTEELVRGWMG----	RVFNQVYV-----	435	
QY 366	---FYNNELHAYLASCGVDGVKYDVQNIITETLAGAGGVSILTRSYHNHAEASIASNTT	-----	420	
DB 436	IGEARHLNLYIELKDSMLTESDVF-----	HCIM-----	482	
QY 421	-----DAGCIACQCHNTDGLYSAKQATAYRASDDFPRDPASHTIHSSVAYNSLFL	-----	472	
DB 483	KHSLIVHGSGGMLGPRNDMSASOCKRSLICKMSDFPRD-----	YKPRLLII-----	532	
QY 473	GEFMQPDWMFHSFLPADYHAAARAGISGCPRIYSDKFGNNFOLLKLYLPDQSVYLRAG	-----	532	
DB 533	LKLMHAD-----	KSLKPRDQFGEEMKLOYI-----	571	
QY 533	LPGRPTDLSLFVDPARDRTSLKILWNLKSGSVGVFNCQAGACMKLEKTRLIHDTSPGT	-----	592	
DB 572	PQCSIN-----	LRVLIHQCS-----	598	
QY 593	LTAASVCAVDLLITQYAGAEWLGDFT-----	YVAYSGEVIRLPRGVSTIPVTLKYLEF	645	
DB 599	LNLEEV-----	LSFANSGISIEMLPSTIGNLKELRYADLTINCGRLIDNG-----	VLKLYKL 648	
QY 646	ELHFHCPIQEIASISFAAIGLLDMENNQGAQVEYEH-----	NRAATKT-----	IA 692	
DB 649	EELMYRGGRIQKALISFTDNCNENAKERSKULSTALEFEFFKNNQOPKMSFENLERKIS	-----	708	

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,538
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Mallen, John W. III
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 188881A
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-232-538-14

Query Match 2.7%; Score 109.5; DB 1; Length 780;
Best Local Similarity 17.7%; Pred. No. 0.017;

Matches 111; Conservative 82; Mismatches 211; Indels 223; Gaps 25;

QY 86 CGRDVPLETQFMLESKESETDGENSPITVYLLPILLEGQFRAVILQNDKNEIICLESQ 145
DB 252 CTATPPLNTFRQMTWSYRDEKNKRAS-----VRRRIDQSNHANIFYSVLTI 298
QY 146 DNAVETQGLHMYMNGATNPEVINOAYKAVEKNMOTFLHREKRLPSCIDMFGMCTWD 205
DB 299 DKMOKNDKGLTYCVRSGPS-FKSVNTSVHIYDKAFITYNHRKQVLE----- 345
QY 206 AFYTDVTAEGVEEGLKLSOGGTP-PRFLIIDGWOQIENKA-----KATECL 253
DB 346 -----TVAGKRSYRLSMKYKAFPSPEVWMLKDGLPATEKSARYLGRGYSLIKDYTE-- 397
QY 254 VOEGAQFATRLTGKENTKFOKLIQNNQMSGLKHLVHGAKOHNNKYNVYWHALAGYNG 313
DB 398 --EDAGNTYLLSTIKQSNVF-----KNLT--ATLIVNKPQIYEKAVSSF-- 438
QY 314 GVKPA--ATGMEHYDTALAYPVQSPGV-----LGNQPDIVMDSLAV 352
DB 439 -PDPAIPLGSRQILCTAGIQRPTIKMFHPCNNHSEARCDPFSNNDESFILADSN 497
QY 353 HG-----LGLVHKKYFNFNELHAYL--ASCGVDGVKVDY-QNIETILGAG----- 396
DB 498 MGNRIESTITORMALIEGK-----NKMASPLVYADSRISGIYICIASNMKVGIVGRNISFY 551
QY 397 ---HGGVRSVLS-----TRSYHIAL----- 411
DB 552 ITDVPNGFHVLEKMPTEGEDKLKSCVNFELVRYDTWILLRTVNNRTHMYSISKQMAI 611
QY 412 --EASISNFT-----DNQCLACQCHNT--DGLYSAKQAIIVRASDEFYRDPASHT 459
DB 612 TRKHSITLNLIMNVSLODSGTACARANNYTGEBEILQKKEITIRQOEAPYLLRNLSDDT 671
QY 460 IHISVAANSLFLGEPQPDMDHSLHPADYHAARAIGCGPIYVSDPKGNHNDLKL 519
DB 672 VAISSTLTDCHANGVPEPQITWFKNNH-----IQDEPG----- 706
QY 520 KLVLPDGSVLRALQGRPTRDSLEVDPAKDRSTSLIKIWNLNKCSGVAVFNCQAGACMKI 579
DB 707 -IILGPGS-----STLFIERYTEDE-----GYVHCKAT----- 734
QY 580 EKKTRIHDSPTGLTASVCAVDVLT 606
DB 735 NOKGSVSSATLVQGTSDKSNLELIT 761

RESULT 7
US-08-786-164-14

Sequence 14, Application US/08786164
Patent No. 5861484
GENERAL INFORMATION:
APPLICANT: THOMAS, KENNETH A.
APPLICANT: KENDALL, RICHARD L.
TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL
CELL GROWTH FACTOR
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Microsoft Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/786,164
FILING DATE: 21-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mark Hand, J
REGISTRATION NUMBER: 36,545
REFERENCE/DOCKET NUMBER: 18888DA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-594-3905
TELEFAX: 908-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-786-164-14

Query Match 2.7%; Score 109.5; DB 2; Length 780;

Best Local Similarity 17.7%; Pred. No. 0.017;

Matches 111; Conservative 82; Mismatches 211; Indels 223; Gaps 25;

QY 86 CGRDVPLETQFMLESKESETDGENSPITVYLLPILLEGQFRAVILQNDKNEIICLESQ 145
DB 252 CTATPPLNTFRQMTWSYRDEKNKRAS-----VRRRIDQSNHANIFYSVLTI 298
QY 146 DNAVETQGLHMYMNGATNPEVINOAYKAVEKNMOTFLHREKRLPSCIDMFGMCTWD 205
DB 299 DKMOKNDKGLTYCVRSGPS-FKSVNTSVHIYDKAFITYNHRKQVLE----- 345
QY 206 AFYTDVTAEGVEEGLKLSOGGTP-PRFLIIDGWOQIENKA-----KATECL 253
DB 346 -----TVAGKRSYRLSMKYKAFPSPEVWMLKDGLPATEKSARYLGRGYSLIKDYTE-- 397
QY 254 VOEGAQFATRLTGKENTKFOKLIQNNQMSGLKHLVHGAKOHNNKYNVYWHALAGYNG 313
DB 398 --EDAGNTYLLSTIKQSNVF-----KNLT--ATLIVNKPQIYEKAVSSF-- 438
QY 314 GVKPA--ATGMEHYDTALAYPVQSPGV-----LGNQPDIVMDSLAV 352
DB 439 -PDPAIPLGSRQILCTAGIQRPTIKMFHPCNNHSEARCDPFSNNDESFILADSN 497
QY 353 HG-----LGLVHKKYFNFNELHAYL--ASCGVDGVKVDY-QNIETILGAG----- 396
DB 498 MGNRIESTITORMALIEGK-----NKMASPLVYADSRISGIYICIASNMKVGIVGRNISFY 551


```

NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-078810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 125:
SEQUENCE CHARACTERISTICS:
LENGTH: 1817 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
FEATURE:
NAME/KEY:
LOCATION: 1..1817
OTHER INFORMATION: /note= "RG2S deduced sequence"
US-09-004-838-125

```

```

Query Match      2.6%; Score 102; DB 4; Length 1817;
Best Local Similarity 18.1%; Pred. No. 0.56;
Matches 141; Conservative 105; Mismatches 266; Indels 268; Gaps 37;

QY 58 FPMGLLEGIRFMC--FRFKL-----MMTQRMGTCGHD-----VPLETQFMLE 100
DB 89 FPIDVI-----TCCSLRIHKLGGKAFKITEQIESLTRQLSLISWTDPPVLAGR---VG 139
QY 101 SKESPTDENSEPIIYTVLLPLLEGQFRAVLQGNPKNEIEICLESQDNVAVETDQGLHWYV 160
DB 140 SMNASTSASSSD-----FPSREKFTQALK-----ALENNQGFHWAL 178
QY 161 --HAGTNEFEVYNQAVKV--EKHMOTFLHRE--KKRLPSCG-----DMFGMCTDAFYT 209
DB 179 CGMGVGVGKTRMMQRLKKAEEKLFENYIVRAVIGKTDPAIOBAIADYLS--IQLNERTK 237
QY 210 DVTAGVEEGLKLSQGGTPPRFLIIDGWOQIE-----NKAKDAECIVQESAO 259
DB 238 PARADKLEMEFKKNSDGKTFELYLDVWQVLVDIEDIGLSPFNQGVDFVLLTSRDSQ 297
QY 260 FAYR-----LTGIKENTKFOKKLQNNP-----QMSGIKHLVHGAK 294
DB 298 VCTWAGVEANSIINVGLLTEAQAQSLFQGFVETSEPELOKIGEDIVKCCCLPAIAIKTMA 357
QY 295 QHNANKYVYVHALAGYGVGKPRATGMEHTALAYP-----VQSP----- 336
DB 358 CTKRKRKRDAM-----KDALSRIEHYDINHVAPKVETSYHNLQEEETKSTFLMC 407
QY 337 GVLGNQPIDVMSLAVHGLGLVHPKRVENFYNELHAVLASGVGVAVDVONIIETLGAG 396
DB 408 GLFPEDPDIPIBELMKRYOMGLKIDRYVTT--REARTRLNTC-----IERLVOT----- 454

```

```

QY 397 HGRVSLTRSYHHALEASIASNFTDNGCIACMCHNTDGLYSAKOTAIVRASDDFYPRDPA 456
DB 455 -----NLTIESDDVGCY-----KMHDLVRA-----FVLGMF 480
QY 457 SHTIHSSAVANSLEFLGFEMQPDMDMFSLHPADYHAAANAIGCPIIYSDKGNANFD 516
DB 481 SEVENASTIVNHGNM-----PEW-----TENDITDSCRRIISLTCKSMKFPDGRFP 526
QY 517 -LLKKILVLPDGSVYLRQAQ--LPGRPTRDSLFPD-----PARDRTSL-LKTNMLNKCSCG 564
DB 527 NMLIKLHMGKSLRFPDPEYEGMEKHLVICYDMKKYPLPLARCSTNNIKVLHLCSCS- 585
QY 565 VGVFNCQAGWCCKIEKTRIHDTSPGTLTASVCASDVDLITQYAGA-----EWLGDTI-- 618
DB 586 -LKMFDSCS-----CIGNLSNLEVLSFANSRIEMLPSTVRN 619
QY 619 -----VYVRSGEVIRLPLKGVSIPTVLKYLEFELFHCPCPIQELNPSIFAIGLID----- 669
DB 620 LKKRLDLRFGDGRLEQGV-LKSLVLEEFYIGN-----ASGFIDNCN 664
QY 670 -----MFTGGAVEQVEIHNRAATR--TIALSVRGGRFRGVYSSORPLKCV 713
DB 665 EMARSDNLSALEFAFNNKAEVKNNMFENLERKISVGRSFGQNTIMSSHYENMLQLY 724

```

```

RESULT 14
US-07-852-260-2
Sequence 2, Application US/07852260
Patent No. 5525715
GENERAL INFORMATION:
APPLICANT: Racanietello, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/852,260
FILING DATE: 19920619
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-B-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-852-260-2

```

```

Query Match      2.6%; Score 102; DB 1; Length 2206;
Best Local Similarity 21.0%; Pred. No. 0.8;
Matches 122; Conservative 67; Mismatches 207; Indels 184; Gaps 33;

```

```

0Y      10 NDGLVNH -GKTLITGVPRNVLTTC-----SR-----GLV 40
           |||:::|
Db      1660 NDGLVNTSKRYPMYTPRGAVTEQGYLMLGGQTARILMTNEPPRACOGCGVITTCGY 1719
           |||:::|
0Y      41 TGAIVGATASHSHSLHFVFWGVLEGRFCSCRRFKLMWTORMTGCGRDVLTEQFMILE 100
           |||:::|
Db      1720 IGMHVGNGSH-----GFAAL-----KRYPLOSQ-----EIOMNR-P 1753
           |||:::|
0Y      101 SKSEETDEMSPIITYLVLPLE-GCFRAVLGNQNKNELEICLESQDNAAVEFDOLHMY 159
           |||:::|
Db      1754 SKEA----GYPLIMAFPTKTLEPSAFHYFVFBEVKEPA---LYKNDRPLRTIDEBEALFS 1805
           |||:::|
0Y      219 GLKSL--SOGGTPRPRLIIDDGMOOIENK---AKDATECLVQBAGFAIRLGISEKENT 271
           |||:::|
Db      1851 GLEBLIDSTSGAYP--YVAGKKRKRIJLKOTFRDTCEMORLLDAYGINLP- VTVVKDEL 1907
           |||:::|
0Y      272 KFOAKLONNEMSGSLKLVLVCAKHNNKVNYVMHALAGWGCVKPRAATGMEHYDTALAY 331
           |||:::|
Db      1908 RSKTKVE-----QCKSLILASLSINDY-----AMRAPFGNLTAAF---H----- 1944
           |||:::|
0Y      332 PVQSPGV-----IGNOPDIWDSILA VHGIQLVHPKKVFNF-YNELHAVLASGVVDGVKD 385
           |||:::|
Db      1945 --RNPGVVTGSVAVCDDDLFWSKIPV-----LMEEKLFAFDYTGYXASLSPAMEFALKMV 1997
           |||:::|
0Y      386 VQNIIEFLGAGHGGRVS-----LTRSYHHALBASIASNFLDNCIACMCNIDGLTSAKOT 441
           |||:::|
Db      1998 LEKI-----GFGRVDYIDLHNS-HHLTKNKI---YCVKGGMBSGCGSTIFNSMJNN 2047
           |||:::|
0Y      442 AIVRA-----SDDF--YPRD-----PASHTIH 461
           |||:::|
Db      2048 LIIRLLLTKYKGIDLDHLMKIAVGDVDVIA SHEVDASILAQSQRDYGLTWTPADKSAFT 2107
           |||:::|
0Y      462 ISSVAY-NSLFGEFMOPMD--MFISLRPADYHAAR 497
           |||:::|
Db      2108 FEYTWENVTFLKRFFRADERKYPEPLIHVPMPKEIHESIR 2147
           |||:::|

RESULT 15
US-08-461-503-2
; Sequence 2, Application US/08461503
; Patent No. 5834302
; GENERAL INFORMATION:
; APPLICANT: Racaniello, Vincent
; APPLICANT: Tateam, Joanne M.
; APPLICANT: Weeks-Levy, Carolyn L.
; TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,503
; FILING DATE: 5-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 36607-D-PCT-US
; TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: (212) 298-0400
: TELEFAX: (212) 371-0525
: TELEX: 422523 COOP UI
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2206 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
US-08-461-503-2

```

Query Match	2.6%;	Score 102;	DB 2;	Length 2206;
Best Local Similarity	21.0%;	Pred. No. 0.8;		
Matches 122;	Conservative 67;	Mismatches 207;	Indels 184;	Gaps 33

Search completed: June 4, 2002, 09:05:32
Job time: 595 sec

Tue Jun 4 16:39:14 2002

us-09-810-186-1.rai

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 09:03:02 ; Search time 24.15 Seconds

(without alignments)
2984.141 Million cell updates/sec

Title: US-09-810-186-1

Perfect score: 3996
Sequence: 1 MTVPKISVNDGKLVHNGKT.....TFSIPSPDEMYRMSIEIQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3219.5	80.6	773	2	imibition protein
2	3181.5	79.6	773	2	imibition protein
3	2951.5	73.9	757	2	Slp1 protein - bar
4	1915.5	47.9	1170	2	protein F20N2.14
5	1307	32.7	807	2	hypothetical prote
6	1228.5	30.7	357	2	probable seed inh
7	753.5	18.9	371	2	hypothetical prote
8	442.5	11.1	649	2	hypothetical prote
9	126	3.2	657	2	probable rep helic
10	125	3.0	982	2	nitrate reductase
11	119.5	3.0	1715	2	resistance protein
12	117.5	2.9	4544	1	alpha-2-macroglobu
13	115	2.9	928	2	polymorphic membra
14	115	2.9	928	2	polymorphic membra
15	114.5	2.9	1447	2	probable protein g
16	114	2.9	4545	1	alpha-2-macroglobu
17	111	2.8	953	2	non-dependent rec
18	110	2.8	905	2	protein F5A9.14
19	109.5	2.7	1338	2	protein-tyrosine k
20	108	2.7	714	2	penicillin-binding
21	106	2.7	976	2	hypothetical prote
22	105	2.6	1601	2	hypothetical prote
23	104.5	2.6	1127	2	hypothetical prote
24	104	2.6	754	2	hypothetical prote
25	104	2.6	1669	2	quantum nucleotide
26	103.5	2.6	587	2	xanthomonasepsin
27	103	2.6	1408	2	cell surface glyco
28	102.5	2.6	386	2	probable ABC trans
29	102	2.6	452	2	probable amino acil

ALIGNMENTS

30	102	2.6	1503	2	T43166	alpha-2-macroglobu
31	102	2.6	1813	2	T30564	resistance protein
32	102	2.6	2206	1	GNNY4P	genome polypotein
33	102	2.6	2206	2	S03832	genome polypotein
34	102	2.6	2207	1	GNNY5P	genome polypotein
35	101.5	2.5	374	2	B82168	probable alpha-1,6
36	101.5	2.5	675	2	B96542	unknown protein
37	101.5	2.5	1158	2	AF1852	hypothetical prote
38	101.5	2.5	1430	2	D82533	RNA polymerase bet
39	101.5	2.5	1477	2	B43855	high-molecular-wet
40	101	2.5	1820	2	S71853	genome polypotein
41	101	2.5	2209	1	GNNY2P	genome polypotein
42	101	2.5	4466	1	S17231	dynein beta heavy
43	101	2.5	4466	1	S17653	dynein beta heavy
44	100.5	2.5	1171	2	T12956	hypothetical prote
45	100	2.5	949	2	E75352	glycine cleavage s

RESULT 1

imibition protein homolog - Arabidopsis thaliana
N:Alternate names: protein T8H10.120
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46188
R:Benes, V.; Rechmann, S.; Borckova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23014
A:Accession: T46188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <BEN>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Insertions: 64/2; 146/1; 176/2; 192/3; 223/2; 259/2; 300/2; 484/3; 507/2; 552/2; 625/1
A:Note: T8H10.120

Query Match	80.6%	Score 3219.5	DB 2	Length 773
Best Local Similarity	76.6%	Pred. No. 1,3e-240		
Matches 593	Conservative 73	Mismatches 83	Indels 25	Gaps 3
QY	1	MTVPKISVNDGKLVHNGKTILGVPDNNVLTPOGSGRLVGAFAVGATASHKSLHVPW	60	
DB	1	MTISNLSVQNDNLVYOGKITILKIPNNILITLVYTGNGFVSGSFGTGFQSKSLHVPF	60	
QY	61	GVLGELFMCCFRKKLMMWTQRMGTGGRVPLETQFMLESK-ESFTDGENSPIYTVLL	119	
DB	61	GVLGELFMCCFRKKLMMWTQRMGTGGRVPLETQFMLESK-ESFTDGENSPIYTVLL	119	
QY	120	PLEGQRAVLQGNKEKEITICESGNAVEETOGGLHMYMAGTINPEFYINQAVAYEK	179	
DB	120	PLEGQRAVLQGNKEKEITICESGNAVEETOGGLHMYMAGTINPEFYINQAVAYEK	179	
QY	121	PLEGQRAVLQGNKEKEITICESGNAVEETOGGLHMYMAGTINPEFYINQAVAYEK	179	
DB	121	PLEGQRAVLQGNKEKEITICESGNAVEETOGGLHMYMAGTINPEFYINQAVAYEK	179	
QY	180	HMOTFLREKKRRLPSCIDMFGMCTWDAFYDYVAEGVEEGIKLSOGGTPPRLLIDDM	239	
DB	180	HMOTFLREKKRRLPSCIDMFGMCTWDAFYDYVAEGVEEGIKLSOGGTPPRLLIDDM	239	
QY	181	HMOTFLREKKRRLPSCIDMFGMCTWDAFYDYVAEGVEEGIKLSOGGTPPRLLIDDM	239	
DB	181	HMOTFLREKKRRLPSCIDMFGMCTWDAFYDYVAEGVEEGIKLSOGGTPPRLLIDDM	239	
QY	240	QOIEENKAKDTECLVDGQAFATRLGIEKNTFFOKLONNEQMSGLKHLVHAKOHNV	299	
DB	240	QOIEENKAKDTECLVDGQAFATRLGIEKNTFFOKLONNEQMSGLKHLVHAKOHNV	299	
QY	241	QOIEENKAKDTECLVDGQAFATRLGIEKNTFFOKLONNEQMSGLKHLVHAKOHNV	299	
DB	241	QOIEENKAKDTECLVDGQAFATRLGIEKNTFFOKLONNEQMSGLKHLVHAKOHNV	299	
QY	300	KNNYVHVALAGYGVKPAATGMEHYDTALAYPVQSGVIGNPDIVMSLVHGLVH	359	
DB	300	KNNYVHVALAGYGVKPAATGMEHYDTALAYPVQSGVIGNPDIVMSLVHGLVH	359	
QY	300	KNNYVHVALAGYGVKPAATGMEHYDTALAYPVQSGVIGNPDIVMSLVHGLVH	359	
DB	300	KNNYVHVALAGYGVKPAATGMEHYDTALAYPVQSGVIGNPDIVMSLVHGLVH	359	
QY	360	PKVFNFNELHAYLASCGVDGVKVDVQNIITLGAHGGRVSLFRSYHHALEASIANF	419	
DB	360	PKVFNFNELHAYLASCGVDGVKVDVQNIITLGAHGGRVSLFRSYHHALEASIANF	419	

Db 360 PKRVFNENHLSYLASCGIDGVKVDVONIETLGAIGGRVSLTSSYOOALFASIAIRNF 419
QY 420 TDNGCIACMCHNTDGLYSAKQTAIVRASDDFPRDPASHTHISSVAANSFLFGEFMOPD 479
Db 420 TDNGCIACMCHNTDGLYSAKQTAIVRASDDFPRDPASHTHISSVAANSFLFGEFMOPD 479
QY 480 WDMFSLHPAADYHAARAAGCGPIYVSDKPGNHNDLKLKLVLPDGSVLRRAOLGPRPTD 539
Db 480 WDMFSLHPAEYHAARAAGCAIYVSDKPGNHNDLKLKLVLPDGSVLRRAOLGPRPTD 539
QY 540 DSLEFVDARPTSLKTIWNLKCSGVYVFNCOGAGMCKIEKTRIHDTSPGTLTASVCA 599
Db 540 DSLEFVDARPTSLKTIWNLKCSGVYVFNCOGAGMCKIEKTRIHDTSPGTLTASVCA 599
QY 540 DCLFADPARRGISLTKIWMNMKFTGIVGFCNCOGAGMCKTKKNOIHDTSPGTLTASVCA 599
Db 540 DCLFADPARRGISLTKIWMNMKFTGIVGFCNCOGAGMCKTKKNOIHDTSPGTLTASVCA 599
QY 600 SDVDLITOVAGAEMLDITIVYARSGEVIRLPKGVSIPTVLKYLEFELHPCPIQETIAPS 659
Db 600 SDVDLITOVAGAEMLDITIVYARSGEVIRLPKGVSIPTVLKYLEFELHPCPIQETIAPS 659
QY 660 ISFAAIGLDMFNTGAVOEVEIH-----NRATKTIALSVR 696
Db 660 ISFAAIGLDMFNTGAVOEVEIH-----NRATKTIALSVR 696
QY 697 GRGRFGVYSSORPLKCVYGAETDFNYDSEGTLTFTSIPVSPBEYRMSIEIQY 750
Db 697 GRGRFGVYSSORPLKCVYGAETDFNYDSEGTLTFTSIPVSPBEYRMSIEIQY 750
QY 720 GCGRFGAYSSORPLKCAVESTETDPTDVAEGLVTLNLPTRREEMFRMVEILY 773
Db 720 GCGRFGAYSSORPLKCAVESTETDPTDVAEGLVTLNLPTRREEMFRMVEILY 773

RESULT 2
S45033
probable imbibition protein - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S45033
R:Fujikura, Y.; Karszen, C. K.
submitted to the EMBL Data Library, May 1994
A:Description: Cauliflower cDNA encoding a putative imbibition protein.
A:Reference number: S45033
A:Accession: S45033
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-765 <FUJ>
A:Cross-references: EMBL:X79330; NID:q488786; PID:q488787

Query Match 79.6%; Score 3181.5; DB 2; Length 765;
Best Local Similarity 76.5%; Pred. No. 1,1e-237;
Matches 587; Conservative 74; Mismatches 87; Indels 19; Gaps 3;
QY 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTPGSGRGLVTAEGVATASHSKSLHVEPM 60
Db 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTPGSGRGLVTAEGVATASHSKSLHVEPM 60
QY 61 GVLGELRFMCCEFRKLMWMTQRMGTGGRVPLETOFMLEIESKETDGENSPIIYVLLP 120
Db 61 GVLGELRFMCCEFRKLMWMTQRMGTGGRVPLETOFMLEIESKETDGENSPIIYVLLP 120
QY 121 LLEGGFRALVLOGNDKNEIEICLESQDNVVEDDGLHMYMAGTPEFVINOAKVAKKH 180
Db 121 LLEGGFRALVLOGNDKNEIEICLESQDNVVEDDGLHMYMAGTPEFVINOAKVAKKH 180
QY 180 MGTFFHREKKLPSFVDMGCTWDAFTDVAEVDGSLRSISGGRPRRLIIDGMO 239
Db 180 MGTFFHREKKLPSFVDMGCTWDAFTDVAEVDGSLRSISGGRPRRLIIDGMO 239
QY 241 QLENKAKDTEBCLVODGAFATRLTGIKENTKFOKKLONNEDMSGLKHLVHGAQHNVK 300
Db 241 QLENKAKDTEBCLVODGAFATRLTGIKENTKFOKKLONNEDMSGLKHLVHGAQHNVK 300
QY 301 NYUWHALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 360
Db 301 NYUWHALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 360
QY 358 KKHNVKVVYVHVALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 358
Db 358 KKHNVKVVYVHVALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 358
QY 361 KKVFNENHLSYLASCGIDGVKVDVONIETLGAIGGRVSLTSSYOOALFASIAIRNF 420
Db 361 KKVFNENHLSYLASCGIDGVKVDVONIETLGAIGGRVSLTSSYOOALFASIAIRNF 420

Db 359 KKVFNENHLSYLASCGIDGVKVDVONIETLGAIGGRVSLTSSYOOALFASIAIRNF 418
QY 421 DNGCIACMCHNTDGLYSAKQTAIVRASDDFPRDPASHTHISSVAANSFLFGEFMOPD 480
Db 421 DNGCIACMCHNTDGLYSAKQTAIVRASDDFPRDPASHTHISSVAANSFLFGEFMOPD 480
QY 419 DNGCIACMCHNTDGLYSAKQTAIVRASDDFPRDPASHTHISSVAANSFLFGEFMOPD 478
Db 419 DNGCIACMCHNTDGLYSAKQTAIVRASDDFPRDPASHTHISSVAANSFLFGEFMOPD 478
QY 481 DMEFSLHPAADYHAARAAGCGPIYVSDKPGNHNDLKLKLVLPDGSVLRRAOLGPRPTD 540
Db 481 DMEFSLHPAEYHAARAAGCAIYVSDKPGNHNDLKLKLVLPDGSVLRRAOLGPRPTD 538
QY 479 DMEFSLHPAEYHAARAAGCAIYVSDKPGNHNDLKLKLVLPDGSVLRRAOLGPRPTD 538
Db 479 DMEFSLHPAEYHAARAAGCAIYVSDKPGNHNDLKLKLVLPDGSVLRRAOLGPRPTD 538
QY 541 DSLEFVDARPTSLKTIWNLKCSGVYVFNCOGAGMCKIEKTRIHDTSPGTLTASVCA 600
Db 541 DSLEFVDARPTSLKTIWNLKCSGVYVFNCOGAGMCKIEKTRIHDTSPGTLTASVCA 600
QY 539 DCLFADPARRGISLTKIWMNMKFTGIVGFCNCOGAGMCKTKKNOIHDTSPGTLTASVCA 598
Db 539 DCLFADPARRGISLTKIWMNMKFTGIVGFCNCOGAGMCKTKKNOIHDTSPGTLTASVCA 598
QY 601 SDVDLITOVAGAEMLDITIVYARSGEVIRLPKGVSIPTVLKYLEFELHPCPIQETIAPS 660
Db 601 SDVDLITOVAGAEMLDITIVYARSGEVIRLPKGVSIPTVLKYLEFELHPCPIQETIAPS 658
QY 599 DADLITSEVAGDPMGDSDIYAYKSGELVRLPKGASIPPLKYLEFELHPCPIQETIAPS 658
Db 599 DADLITSEVAGDPMGDSDIYAYKSGELVRLPKGASIPPLKYLEFELHPCPIQETIAPS 658
QY 661 ISFAAIGLDMFNTGAVOEVEIH-----NRATKTIALSVRGRGRGV 703
Db 661 ISFAAIGLDMFNTGAVOEVEIH-----NRATKTIALSVRGRGRGV 703
QY 659 SFAPIGLDMFNSSGAISQMEINVTYDEKPELSSSVSENNSPALSLGRCGRFGA 718
Db 659 SFAPIGLDMFNSSGAISQMEINVTYDEKPELSSSVSENNSPALSLGRCGRFGA 718
QY 704 YSSORPLKCVYGAETDFNYDSEGTLTFTSIPVSPBEYRMSIEIQY 750
Db 704 YSSORPLKCVYGAETDFNYDSEGTLTFTSIPVSPBEYRMSIEIQY 750
QY 719 YSSORPLKCVYGAETDFNYDSEGTLTFTSIPVSPBEYRMSIEIQY 750
Db 719 YSSORPLKCVYGAETDFNYDSEGTLTFTSIPVSPBEYRMSIEIQY 750

RESULT 3
S27762
Sip1 protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27762
R:Heck, G.R.; Dorsett, C.; Ho, T.H.
submitted to the EMBL Data Library, February 1992
A:Description: Cloning and characterization of a gene, Sip1, associated with seed imb
A:Reference number: S27762
A:Accession: S27762
A:Molecule type: DNA
A:Residues: 1-757 <HEC>
A:Cross-references: EMBL:M77475; NID:q167099; PID:q167100
C:Genetics:
A:Introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match 73.9%; Score 2951.5; DB 2; Length 757;
Best Local Similarity 71.1%; Pred. No. 6.6e-220;
Matches 540; Conservative 92; Mismatches 116; Indels 11; Gaps 5;
QY 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTPGSGRGLVTAEGVATASHSKSLHVEPM 60
Db 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTPGSGRGLVTAEGVATASHSKSLHVEPM 60
QY 61 GVLGELRFMCCEFRKLMWMTQRMGTGGRVPLETOFMLEIESKETDGENSPIIYVLLP 113
Db 61 GVLGELRFMCCEFRKLMWMTQRMGTGGRVPLETOFMLEIESKETDGENSPIIYVLLP 113
QY 114 IYUWPLLEGGFRALVLOGNDKNEIEICLESQDNVVEDDGLHMYMAGTPEFVINOAKVAKKH 173
Db 114 IYUWPLLEGGFRALVLOGNDKNEIEICLESQDNVVEDDGLHMYMAGTPEFVINOAKVAKKH 173
QY 120 VILVMDLLEGGFRALVLOGNDKNEIEICLESQDNVVEDDGLHMYMAGTPEFVINOAKVAKKH 179
Db 120 VILVMDLLEGGFRALVLOGNDKNEIEICLESQDNVVEDDGLHMYMAGTPEFVINOAKVAKKH 179
QY 174 VKAVERKMTQFFHREKKLPSFVDMGCTWDAFTDVAEVDGSLRSISGGRPRRLIIDGMO 233
Db 174 VKAVERKMTQFFHREKKLPSFVDMGCTWDAFTDVAEVDGSLRSISGGRPRRLIIDGMO 233
QY 234 IIDDGMOIENKAKDTEBCLVODGAFATRLTGIKENTKFOKKLONNEDMSGLKHLVHGA 293
Db 234 IIDDGMOIENKAKDTEBCLVODGAFATRLTGIKENTKFOKKLONNEDMSGLKHLVHGA 293
QY 240 IIDDGMOIENKAKDTEBCLVODGAFATRLTGIKENTKFOKKLONNEDMSGLKHLVHGA 298
Db 240 IIDDGMOIENKAKDTEBCLVODGAFATRLTGIKENTKFOKKLONNEDMSGLKHLVHGA 298
QY 294 KOHNHVNVYVHVALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 353
Db 294 KOHNHVNVYVHVALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 353
QY 299 KKHNVKVVYVHVALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 358
Db 299 KKHNVKVVYVHVALAGVGVKPRATGMEHDTALAYVOSPGVLGNOPDIYVMSLAVHGLVNP 358

QY	111	SPYLYLLEGGORAVLQGNKDNKEIETCESSDNVAVEIDQGLAHVWYAGNAGNEPEVI	170
Db	151	----YAAITPLEGAFRASITPEKQNVLTCESSSTYKESSEKSTAIYHICCNPLNMA	206
QY	171	NOAKVAKEMQOTELHREKRLPSCLDIMGWCMTWDAFYTYVTAEGVEGILKSLSQGTPP	230
Db	207	KEAFSALRYAHMNTFKLEEKLPIKYDRGCMCTWACYLTVDPATIMWGKEPEDGVCY	266
QY	231	RFLLIIDGMOOI-----ENKAKDTELCVQEGQFATRLTGIKENKF---OKTQNNQOM	283
Db	267	KFVLIIDGMOISINFDGDELKDA-ENMLVGGQDMARTLITSECKKFFNNYEESLSGDDV	325
QY	284	SG-----LKLHVGAKOHNNYKNVYMHALGVMYGWGPAAATGMEHDTALAYPVQSPG	337
Db	326	SGSSMAAFYTDL---RLRFESLDDIYVYHALGANGVAPRP- -YMDKAKVAPFELS	380
QY	338	VLGNPQDIYVMSDLAVHGLGLVHPKVFNFYNELHAYLASCGVGVKVDYQNIIEITLGAH	397
Db	381	LGATMADLADKVVLEGVIGIHPKSAHEEYDSHSLTASVGTGAKIDVFPOTLESLAEH	440
QY	398	GGARSLTFRSHHMLEASIANFTDNQCTACMOCHNTDGLYSA- -KQAIYRASDDPYPRP-	455
Db	441	GGARVELKAAIYDGLTESMIKENGCTVYIASMOQCNFFFLAKKJSTIGRVGDDPWQDPY	500
QY	456	-----ASHHTISISVAVNSLFLGEGMOPDMDMFSLHRAADYHAAARATGCGPIY	505
Db	501	GDPQGVWLGCVHHIICS---YNSIMMGQIDOPDMDMFQSDIYCAEYHAASRALGCGPY	557
QY	506	VSDKPG--NNHFDLKLUYLPDGSVYLRADLPGRPTDSIFYVPAADRSTILKIMNLNCS	563
Db	558	LSDHGASAHNFEDLIRKLAFFDDTIPRCVYHALPYTDSLEFKNPLDEKESLITFENKFG	617
QY	564	GVVGVENCQAGMCKIKKTRIDTSPGLTASVCASDV--DLITQVAGAE--WLGDITV	619
Db	618	GVITGFNCQAGMSPBEHRKGYKECYTYSGVYHVSQDIEMQNEAAGSQVITYGDYLV	677
QY	620	YARSGEVIRL- -PKGVSIPVTLKVLVELEFHEHCPICQIAPS--ISPAIGLDMTNGAV	677
Db	678	YKQOSEEILFNNSKSEAMKITLESFAFDLLSEFVYTELVSQGVRAPIGLIMFNCGTV	737
QY	678	FOVELHNAAKTIALSVRGSRGVYSSORPLKCVVGALENDVNDSTGLTTPSIP	735
Db	738	ODMKY--TGDNSTRVYCKGGRRMAIYSSAPKCYLNDKEAEFKMEETGLSLFVP	792

```

RESULT      6
T09530
probable seed inhibition protein - chickpea (fragment)
C:/Species: Cicer arietinum (chickpea, garbanzo)
C:/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:/Accession: T09530
R:/Cervantes, E.
submitted to the EMBL Data Library, February 1996
A:/Reference number: Z16718
A:/Accession: T09530
A:/Status: preliminary; translated from GB/EMBL/DBJ
A:/Molecule type: mRNA
A:/Residues: 1-357 <CER>
A:/Cross-references: EMBL:X95875; NID:q1212811
A:/Experimental source: germinating seed
C:/Genetics:
A:/Gene: slp

```

	Query Match	Similarity	30.7%	Score 1228.5	DB 2	Length 357
	Best Local	Similarity	68.0%	Pred. No. 4e-87		
Matches	247	Conservative	24	Mismatches	75	Indels 17; Gaps 6;
OY	172	QAVKAVKHMOTFLIREKRRLPSCLDWRCWMDAFYDTVAEGVGGKSLSSGGTPPR	231			
DB	4	QAVKAVKHMOTFLIREKRKRPVSLDFWRCWMDAFYDTVAEGVGGKSLSSGGTPPR	63			
OY	232	FLIIDGGWQILENKAKDATECTLVQSGAOPATRLTG--IKENTKRFQKKLONNEMQSLKHL	289			

```

Db      64 FLITMVGNRKLRKQKSGSCVYQDEBQOL---LLGWMIDANSKKKNQOBNQDGLGKHL 120
QY      290 VHGAKOHNNHKNYUVMHLAGVYGVPKPAATGMEHTDALATPVOSPGVIGNOPDIVDS 349
Db      121 VDGKKNHNNYDUVMHLAGVYGVPKPAATGMEHTDALATSAYTRST-WMOPDIVDS 179
QY      350 LAVHGLGLVHPRKKVFENYNELHAYTLASCGVDGKVDYOVOMIETLGAGHGGRVSLRSYHN 409
Db      180 LAVHGLRLHYHPRG-FISTNETHAYTLASCGVDGKVDYOVOTLLRPLYLD-----TYVESRLH 233
QY      410 A-----LEASIASNFTNGCIACMCHNTDGLYSAKOTALIVRASDDFYPDPASHTHTISS 464
Db      234 AAILMRLRLPLLYTLVLLNNGCIACMCHNTDGLYSAKOTALIVRASDDFYPDPASHTHTISS 293
QY      465 VAVNSLFLGEMQPDWMFHSLSHPADYHAAAAATGGCPTIYSDKRGNNPFDLLKVIPL 524
Db      294 VAVNSLFLGEMQPDWMFHSLSHPADYHAAAAATGGCQPMASYISQATTILLFLSLSYLA 353
QY      525 DGS 527
Db      354 DGS 356

```

RESULT 7
T01717
Hypothetical protein A_I6002N01.5 - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C.Accession: T01717
R.Scheet, P.: Magg, L.
submitted to the EMBL Data Library, June 1997
A.Description: The sequence of A. thaliana I6002N01.
A.Reference number: Z14407
A.Accession: T01717
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-371 <SCH>
A.Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A.Experimental source: cultivar Columbia
C.Genetics:
A.Map position: 4
A.Introns: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 284/1; 295/1; 333/2; 349/2
A.Note: A_I6002N01.5

[illegible]

QY 355 LCLVHPKRVENYNELHNLAVLASCVDGVKVVDQNIIETLGAGHGGRVSLTRSYHHALERS 414
| | | | |
Db 258 L-----GGGVLAKKKYHQALEAS 275

QY 415 IASNFTDNGCIACMKCHNTDGLYSAKOTAIVRASDDFYPRDPASHHTIHSVSANSLFLGE 474
| : : : : :
Db 276 ISRNPPAG-----KHHDLGYCSHS---QROLFHWHRDDPASRHTIHASVAYNLTFLFGE 325

QY 475 FMQPMDMFHSLHPADYHAARAIAGCGPIYVSKDPGNHNPELLKKLVLPSSYL 529
||||||| : : ||||| ||||| |||||
Db 326 FMPQPMDDISS-----SWRMCHLCOVITISDKPGGHDFMLKRLVLDGSIL 371

RESULT 8
D90436
hypothetical protein SS03127 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: D90436
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweyer, M.J.; Chan,
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: D90436
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1'649 <KUR>
A:Cross-references: GB:AE006641; NID:g13816550; PIDN:AAK43227.1; GSPDB:GN00155
Genetics:
Gene: SS03127

```

Db      480  PIYIDRHPERNTIELRNAVLNCEVIRVDEPAITEDDLFKDPLRRRY-LTKIKGKVK 538
QY      562  GCGVGVEN 570
           : 11
Db      539  GYNATAFFN 547

RESULT  9
B11367
probable rep helicase, single-stranded DNA-dependent ATPase (rep) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: B11367
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
erson, J.; Kalk, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A11250; MUID:98332770
A:Accession: B11367
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-657 <COL>
A:Cross-references: GB:AE001194; GB:AE000520; NID:g332355; PIDN:AAC65097.1; PID:g3322
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0102
C:Superfamily: helicase II

```

Query Match	3.2%	Score 126;	DB 2;	Length 657;
Best Local Similarity	21.2%	Pred. No. 0.19;		
Matches 123;	Conservative 68;	Mismatches 220;	Indels 170;	Gaps 29;

	Query Match	11.1%	Score 442.5;	DB 2;	Length 649;	
	Best Local Similarity	26.6%;	Pred. No. 6.1e-26;			
	Matches 146;	Conservative 71;	Mismatches 189;	Indels 143;	Gaps 19;	
QY	51 HSKSLHVPKCVLEGLRFMCCEFRFKLMMTQMGCRDVPLETOFMILIESKESETDGEN	110				
Dd	113 HEKYIRTP-----CWLDPVF-----NNIPITYFALVKSGNS-----	145				
QY	111 SPIIYTVLLPLEGGFRAVLQGNDKNEIEICLESGDNAVETOGCLHMVMHGAT--NPPE	168				
Dd	146 ----YEAFELISNNVAVTLFGD-----SVRIYTFN---IDELKRSSFLSIGTDNRPK	193				
QY	169 VINAQVAKEVHKMOTFLREKKRLP-SCLDMFGWCTWDAAFTT-DYTAAGVERGLKLSQG	226				
Dd	194 ALENAINIASKETFEFLERKEGFPDKVMNGICGSMAAFLTKDLNENELIKVVGIIER	253				
QY	227 GTPREFLIIDGQNOIEKKADATECLVQEGAQFATRLTGILENTKFQOKLONNFQMSG	286				
Dd	254 GLRLNWVIIDGQMDQNND-----RAIRSLNPDKKKFPN-----GF	289				
QY	287 KHLVHGAKOHNNKVNYMHLAGVGGVKPPAANGMEHYDALAPVQSPGVLGNQPDIV	346				
Dd	290 KNTPRAIKS-LGVATYVGLIMHAINAHWGMSGE-----L	321				
QY	347 MDSLAVHGLGLVHKRVFNFYNELHAYLASGV-----DGVKVDVON	388				
Dd	322 MKSLNVNGY-----FTNFLNSIYVPSNLBEALIGFYKAFDGNIILRDDLKAVKDQW	371				
QY	389 IIEF-----LGAGHGGRVSLTRSYTHALEASIASNFTDNCIACMCNHTDGLYSAKOTA	442				
Dd	372 VIHAIVDSFPIGLA-----SRNIQLAQXSVGD-----VINCMSNPNENCYCYFSN	419				
QY	443 IVRASDFYPDPDPAHSHTIHSSVAYNSLFLGEEPMOPMDMFHSHLPADYAARAATGCC	502				
Dd	420 VMRSIDIVPRWKDGTKLIMFNANSLSTSHIYYRPDYDMFSYDPYAKVHIIVAFVEFGG	479				
QY	503 PIYVSDK-PGHNHFDDLKLVLPDQGSVYLRAQLPGRPRTDSLFVDPARBRSTSLKIWNLNK	561				

QY	208	YTDVAEEVGEKLSOGGTPPRELIDD-----	CMQO-----	241
Db	56	FTNKAHHEMSERIKALT--GKPLRNTTVSTHBAGVITLREHIVHLGKRKKEFSYDENDK		113
QY	242	---IENKAKDA-----TECLVOEGAOFATRLTGIEKNEKFOFKLONNOMSGLKHLY		290
Db	114	RALLIREAKKEVHLLPEVLDNCGV-----STLSAIKMQKNKLELOHKRER--ALMHEY		164
QY	291	HGAKOHNVKQVYVWHALAGYWGCVKPAATGMEHYDTALAYVPOSPGVLGNOPIYDLSL		350
Db	165	HTALKLFNAVQ-----FDDLIVLPDQ---IFSEYPDIIASAKY		198
QY	351	AVHGGIVLHPKVFNEYN-----ELHAYIASCGVDGKVDKVDQVNIETLAGHGRSLT		404
Db	199	SRY-----HYLVBEFDQTSAAQYRLMKMLATQNICVAGDDQDSIYSWRCAGHNDNILLSFE		253
QY	405	RSYHHALEASIASNFETDNGCI-----ACMCNNTGGLYSAKOTALIVRASDD-----FYPR		453
Db	254	KDFPATITITILEQVNRSTGTLTILAAANGVIAHNT-----QREKALIMSGNDGSKPLREITNPE		309
QY	454	DPASHTIHI-----SSVAYNS-----LFIGEFNOQPDWMEHSLHPADYHAARAAY		499
Db	310	TERDEAIFRANTILAEQIRASYSFSSFGVLLRTNSFMRITEDVE--LQENIPY---RVS		363
QY	500	GGCPIYVSDKRGNNIFDLKLVLVLPDGSYVLAQOLPGRPTRDSLFVDPARDRTSLK--TWN		538
Db	364	GGMGFEF--ORKEIKDVLSTLYRIVSNPDDVNLRLIINPRRI-----GKTLHLIVSDIAN		417
QY	559	LNKSGVGVVENCOGAGCMCKTEKTRLIHDTSPGTLTASVACSDVLTITVOAGAEMLDDTI		618
Db	418	TQOQS-----VENALN-----QIINKTHAVDLKESHRTA--VENFQLLTOARTHLSSKNL		467
QY	619	VYAYRSGEVIRLPGVSIPTVLKLEFE-----LFHFPCPIQELIAPSI-----SFAAIGL		667
Db	468	AYKVR-----KFVEDIOY--FMYLIEOFQKNENHARFKFLQIEHLVESIEHMEQSSHEGSL		521
QY	668	LDMEN-----TGGAVEQVEIHNRAAKTIALSY		695
Db	522	YDYLNRVTLARDNVOKETEGAVSLMTIH---ASKGLEFPV		559

Db 709 VGCYKGDGFKIFRSENFILRLVNTNRE 736

RESULT 12

S02392

alpha-2-macroglobulin receptor precursor - human

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C:Species: Homo sapiens (man)

C:Date: 14-Aug-1998 #sequence-revision 14-Aug-1998 #text-change 22-Jun-1999

C:Accession: S02392; S30027; I37998; A39210; S12538

R:Herz, J.; Hamann, U.; Rognes, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.

EMBO J. 7, 4119-4127, 1988

A:Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein

A:Reference number: S02392; MUID:89210795

A:Accession: S02392

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-4544 <HER>

A:Cross-references: EMBL:X13916; NID:g34338; PIDN:CAA32112.1; PID:g34339

R:Kristensen, T.

submitted to the EMBL Data Library, October 1990

A:Reference number: S30027

A:Accession: S30027

A:Molecule type: mRNA

A:Residues: 3275-3864 <KR>

A:Cross-references: EMBL:X55077

R:Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.

EMBO J. 9, 1769-1776, 1990

A:Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein

A:Reference number: S12538; MUID:90269210

A:Contents: annotation; site of proteolytic cleavage

R:Kuit, H.; Herz, J.; Stanley, K.K.

Biochim. Biophys. Acta 1009, 229-236, 1989

A:Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promote

A:Reference number: I37998; MUID:90089395

A:Accession: I37998

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-11 <RES>

A:Cross-references: EMBL:X15424; NID:g34408; PIDN:CAA3464.1; PID:g34409

R:Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argaves, J.

Biol. Chem. 265, 17401-17404, 1990

A:Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip

A:Reference number: A39210; MUID:91009181

A:Accession: A39210

A:Status: preliminary

A:Molecule type: protein

A:Residues: 150-166;234-238, 'X', 240-245, 'X', 247-252, 'G', 686-695;902-916;1096-1109; 'S', 17

C:Genetics:

A:Gene: GDB:LRP1; APR; LRP; A2MR

A:Cross-references: GDB:119694; OMIM:107770

A:Map position: 12q13.1-12q13.3

C:Complex: The alpha2-macroglobulin receptor complex consists of noncovalently-associated protein (see PIR:A39875).

C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding

C:Keywords: beta-hydroxyaspartagine; beta-hydroxyaspartic acid; calcium binding; glycoprotein

F:1-19/Domain: signal sequence <status predicted <SIG>

F:207-3943/Product: alpha-2-macroglobulin receptor 515k chain #status predicted <S15K>

F:27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:115-148/Domain: EGF homology <EG1>

F:115-188/Domain: EGF homology <EG2>

F:198-239/Domain: LDL receptor WYTD-containing repeat homology <YW01>

F:240-281/Domain: LDL receptor WYTD-containing repeat homology <YW02>

F:293-334/Domain: LDL receptor WYTD-containing repeat homology <YW03>

F:335-378/Domain: LDL receptor WYTD-containing repeat homology <YW04>

F:379-420/Domain: LDL receptor WYTD-containing repeat homology <YW05>

F:421-468/Domain: LDL receptor WYTD-containing repeat homology <YW06>

F:478-519/Domain: EGF homology <EG3>

F:571-613/Domain: LDL receptor WYTD-containing repeat homology <YW07>

F:614-659/Domain: LDL receptor WYTD-containing repeat homology <YW08>

F:660-710/Domain: LDL receptor WYTD-containing repeat homology <YW09>

F:711-752/Domain: LDL receptor WYTD-containing repeat homology <YW10>

F:753-799/Domain: LDL receptor WYTD-containing repeat homology <YW11>

F:807-842/Domain: EGF homology <EG4>

F:854-890/Domain: LDL receptor WYTD-containing repeat homology <YW12>

F:895-931/Domain: LDL receptor WYTD-containing repeat homology <YW13>

F:936-971/Domain: LDL receptor WYTD-containing repeat homology <YW14>

F:976-1011/Domain: LDL receptor WYTD-containing repeat homology <YW15>

F:1015-1051/Domain: LDL receptor WYTD-containing repeat homology <YW16>

F:1062-1097/Domain: LDL receptor WYTD-containing repeat homology <YW17>

F:1104-1140/Domain: LDL receptor WYTD-containing repeat homology <YW18>

F:1145-1182/Domain: LDL receptor WYTD-containing repeat homology <YW19>

F:1185-1221/Domain: EGF homology <EG5>

F:1227-1261/Domain: EGF homology <EG6>

F:1269-1308/Domain: LDL receptor WYTD-containing repeat homology <YW20>

F:1309-1355/Domain: LDL receptor WYTD-containing repeat homology <YW21>

F:1356-1398/Domain: LDL receptor WYTD-containing repeat homology <YW22>

F:1399-1445/Domain: LDL receptor WYTD-containing repeat homology <YW23>

F:1446-1488/Domain: LDL receptor WYTD-containing repeat homology <YW24>

F:1489-1531/Domain: LDL receptor WYTD-containing repeat homology <YW25>

F:1540-1578/Domain: EGF homology <EG7>

F:1583-1626/Domain: LDL receptor WYTD-containing repeat homology <YW26>

F:1627-1669/Domain: LDL receptor WYTD-containing repeat homology <YW27>

F:1670-1713/Domain: LDL receptor WYTD-containing repeat homology <YW28>

F:1714-1753/Domain: LDL receptor WYTD-containing repeat homology <YW29>

F:1754-1796/Domain: LDL receptor WYTD-containing repeat homology <YW30>

F:1797-1846/Domain: LDL receptor WYTD-containing repeat homology <YW31>

F:1850-1886/Domain: EGF homology <EG8>

F:1934-1976/Domain: LDL receptor WYTD-containing repeat homology <YW32>

F:1977-2019/Domain: LDL receptor WYTD-containing repeat homology <YW33>

F:2020-2063/Domain: LDL receptor WYTD-containing repeat homology <YW34>

F:2064-2105/Domain: LDL receptor WYTD-containing repeat homology <YW35>

F:2106-2151/Domain: EGF homology <EG9>

F:2159-2194/Domain: LDL receptor WYTD-containing repeat homology <YW36>

F:2199-2241/Domain: LDL receptor WYTD-containing repeat homology <YW37>

F:2253-2294/Domain: LDL receptor WYTD-containing repeat homology <YW38>

F:2344-2388/Domain: LDL receptor WYTD-containing repeat homology <YW39>

F:2389-2429/Domain: LDL receptor WYTD-containing repeat homology <YW40>

F:2430-2473/Domain: LDL receptor WYTD-containing repeat homology <YW41>

F:2482-2517/Domain: EGF homology <EG10>

F:2524-2561/Domain: LDL receptor WYTD-containing repeat homology <YW42>

F:2566-2600/Domain: LDL receptor WYTD-containing repeat homology <YW43>

F:2605-2639/Domain: LDL receptor WYTD-containing repeat homology <YW44>

F:2652-2688/Domain: LDL receptor WYTD-containing repeat homology <YW45>

F:2696-2739/Domain: LDL receptor WYTD-containing repeat homology <YW46>

F:2734-2769/Domain: LDL receptor WYTD-containing repeat homology <YW47>

F:2774-2812/Domain: LDL receptor WYTD-containing repeat homology <YW48>

F:2818-2853/Domain: LDL receptor WYTD-containing repeat homology <YW49>

F:2858-2897/Domain: LDL receptor WYTD-containing repeat homology <YW50>

F:2904-2939/Domain: LDL receptor WYTD-containing repeat homology <YW51>

F:2944-2980/Domain: EGF homology <EG11>

F:2986-3021/Domain: EGF homology <EG12>

F:3029-3068/Domain: LDL receptor WYTD-containing repeat homology <YW52>

F:3069-3113/Domain: LDL receptor WYTD-containing repeat homology <YW53>

F:3114-3156/Domain: LDL receptor WYTD-containing repeat homology <YW54>

F:3157-3200/Domain: LDL receptor WYTD-containing repeat homology <YW55>

F:3201-3241/Domain: LDL receptor WYTD-containing repeat homology <YW56>

F:3242-3284/Domain: LDL receptor WYTD-containing repeat homology <YW57>

F:3294-3330/Domain: EGF homology <EG13>

F:3334-3369/Domain: LDL receptor WYTD-containing repeat homology <YW58>

F:3374-3408/Domain: LDL receptor WYTD-containing repeat homology <YW59>

F:3413-3448/Domain: LDL receptor WYTD-containing repeat homology <YW60>

F:3453-3489/Domain: LDL receptor WYTD-containing repeat homology <YW61>

F:3494-3531/Domain: LDL receptor WYTD-containing repeat homology <YW62>

F:3536-3570/Domain: LDL receptor WYTD-containing repeat homology <YW63>

F:3575-3609/Domain: LDL receptor WYTD-containing repeat homology <YW64>

F:3613-3647/Domain: LDL receptor WYTD-containing repeat homology <YW65>

F:3654-3690/Domain: LDL receptor WYTD-containing repeat homology <YW66>

F:3695-3731/Domain: LDL receptor WYTD-containing repeat homology <YW67>

F:3731-3776/Domain: LDL receptor WYTD-containing repeat homology <YW68>

F:3785-3822/Domain: EGF homology <EG14>

F:3828-3860/Domain: EGF homology <EG15>

F:3868-3911/Domain: LDL receptor WYTD-containing repeat homology <YW69>

F:3912-3969/Domain: LDL receptor WYTD-containing repeat homology <YW70>

F:3944-4544/Product: alpha-2-macroglobulin receptor 85k chain #status predicted <85k>
 F:3944-4420/Domains: 85k chain extracellular #status predicted <Ext>
 F:3970-4012/Domains: LDL receptor YWTD-containing repeat homology <YW42>
 F:4013-4056/Domains: LDL receptor YWTD-containing repeat homology <YW43>
 F:4057-4099/Domains: LDL receptor YWTD-containing repeat homology <YW44>
 F:4100-4142/Domains: LDL receptor YWTD-containing repeat homology <YW45>
 F:4151-4183/Domains: EGF homology <EG16>
 F:4200-4231/Domains: EGF homology <EG17>
 F:4236-4267/Domains: EGF homology <EG18>
 F:4272-4303/Domains: EGF homology <EG19>
 F:4308-4339/Domains: EGF homology <EG20>
 F:4344-4374/Domains: EGF homology <EG21>
 F:4377-4408/Domains: EGF homology <EG22>
 F:4421-4444/Domains: transmembrane #status predicted <TM>
 F:4445-4544/Domains: Intracellular #status predicted <INT>
 F:166-2998/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F:2998/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F:4075-4125/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 117.5; DB 1; Length 4544;
 Best Local Similarity 20.8%; Pred. No. 16; Mismatches 178; Indels 145; Gaps 25;
 Matches 98; Conservative 51;

OY 71 CFFRLMMWTOGMGTC-----GRDVPLETOFMLESKESEPTDENSEPIITVLLPL 121
 DB 3828 CLRF-----GTCSQLCNNTKGHLCSARNFM---KTHNCKMGSE--YQVLXIA 3873
 OY 122 LEGQRAVLQGNKNEIEICLESQNAVEFD-QGLHM---VY--MHAGTNPPEVINO 173
 DB 3874 DNEIRSLFPGHPHSAEQAFQ-GDESVRIDAMDVHAKGVYWTMTGTITISYSLPPA 3932
 OY 174 VKAV--EKHMQ-----FLHREKKRLPSCLDWFGWCMTWDAFYDVAEGVE-----EG 219
 DB 3933 ADPTTNNRRRDRIDRGVHNLISGLKMPRGIA-IDWAGANVYWDGSDGVLEVAKMGKN 3991
 OY 220 LKSLSGG-TTPRFLIID-----DGM---OQIENKAKAT--ECLVQSGAFATRLT 265
 DB 3992 KRTLLSGMIDEPHAIYVDPRLGTWYSDMGNHPKLETAAMQGLRETLVQONIQMPTGL- 4050
 OY 266 GIKENTKFOKKLQNNQSGKHLVHGAKQHHNKNVYVHNAIAGYNGCVAPRATGMHY 325
 DB 4051 -----AVDYHNERL-----YMDAKLSYVG----- 4070
 OY 326 DFLAAYVQSPGVINOPDIYMDSLAVHGLVHPKVFNFYNEELHAYLASGVGVYVD 385
 DB 4071 -----SIRKNGTDPVYADSKR-----GLSHPSIDVPEDYI-----GVYI 4108
 OY 386 VONITETLGGHGRVSLTSYHAEIASI-----ASNETDNG-----CI-----A 426
 DB 4109 NNRVFKIRHFGHSPVNLGLSHASDVLYHQHKOPEVTNCDRKKCEMLCLLSPSGPV 4168
 OY 427 CMCNTDGLYSAKQTAIVASDPFPPROPASHTIHSSAVYSLPLGEMOP 478
 DB 4169 CTCPCMKRILNCG--TCVPVPSPTPPDAPRGTCMLQCPNGSGCLNARROP 4218

RESULT 13

G65546
 polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strain
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
 C:Accession: G65546
 Nucleic Acids Res. 28, 2311-2314, 2000
 A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A86491; M0ID:20330349
 A:Accession: G65546
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-928 <STO>
 A:Cross-References: GB:BA000008; NID:98978821; PIDN:BA98657.1; GSPDB:GM00142
 A:Experimental source: strain J138

C:Genetics:
 A:Gene: omp_10
 C:Superfamily: Chlamydia pneumoniae polymorphic outer membrane protein G

Query Match 2.9%; Score 115; DB 2; Length 928;
 Best Local Similarity 24.2%; Pred. No. 2.3;
 Matches 107; Conservative 39; Mismatches 145; Indels 152; Gaps 23;

OY 338 VLGNOPDIYMDSLAVHGLVHPKVFNFYNEELHAYLASGVGVYVDVONITETLGGH 397
 DB 271 ISGNQSVTFSGNOAVANGGAIYAKRL-----TLASGGGGISFS-NNIVQSTTAGN 320
 OY 398 GGRVSLTSYHAEIASINFTDNGCIACMCNTDGLYSAKQTAIVASDDFPPDAS 457
 DB 321 GGAISITLAGECSLSAE-AGDITFNG-----NAIVATF--PQTKR 358
 OY 458 HTIHSSVAINSLFLGEGQPMQDMFHSIHPADYHAARAIGGCPRIYSDKPGHNDL 517
 DB 359 NSIDISTAKIT-----NLRAISGHSIF----- 381
 OY 518 LKKLVLPDGSVLRAQLPGRPTRSLFVDP-----ARDRTSLIKINLNLKC-----SG 564
 DB 382 -----PYDEPTANTADSTDL--NLNKADAGNSTDYSG 413
 OY 565 VVGVFNCQAGMCKIEKTRIHDTs---PGLTAS--VCASDVLD---ITQVAGAEWL 614
 DB 414 SI-VFS--GEKLSDEAKVADNLSTLKQPVLTAGNLKRGVTLDTKGFOTAGSSVT 470
 OY 615 GDTIVAYRSGEYIRLPKGVSTPYT-----LVLEELHPQPIQEIAPISFAIG--- 666
 DB 471 MDAGTTLKASTETVTL-TGLSTIPVDSLGEGRKV-----IASAASKVALSGPL 520
 OY 667 LIDMFGTGAVEQVEIHNRAATKTIALSYGRGREGVSSQRLKCVGGAETDENYDSE 726
 DB 521 LLD--NQNAYENHDLCKQDSEFVQLSALGTAT---TTDPAVAPTV-ATPTHHGYQGT 573
 OY 727 TGLT---TFSTVSPDEMYRWS 745
 DB 574 WGMTWVDVDTASTPTKTATLATMT 596

RESULT 14

G81591
 polymorphic membrane protein G family GP0303 [imported] - Chlamydia pneumoniae (s
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 15-Jun-2001
 C:Accession: G81591
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 , C.; Dodson, R.; Gilm, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzebe
 , C.; Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; M0ID:20150255
 A:Accession: G81591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-928 <RBA>
 A:Cross-References: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38160.1; PID:g718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0303
 C:Superfamily: Chlamydia pneumoniae polymorphic outer membrane protein G

Query Match 2.9%; Score 115; DB 2; Length 928;
 Best Local Similarity 24.2%; Pred. No. 2.3;
 Matches 107; Conservative 39; Mismatches 145; Indels 152; Gaps 23;

OY 338 VLGNOPDIYMDSLAVHGLVHPKVFNFYNEELHAYLASGVGVYVDVONITETLGGH 397
 DB 271 ISGNQSVTFSGNOAVANGGAIYAKRL-----TLASGGGGISFS-NNIVQSTTAGN 320
 OY 398 GGRVSLTSYHAEIASINFTDNGCIACMCNTDGLYSAKQTAIVASDDFPPDAS 457


```
Db 321 GGAISIIAGBCSLSAE-ADDITFNG-----NAIVATT-----PQTRK 358
Qy 458 HTIISSVAANSFLGFGKOPDMDHSHLPADYHAARAIGCPITYSDKPGNHFDL 517
Db 359 NSIDIGSTAKIT-----NRAISGHSIF-----381
Qy 518 LKRLVLDGSVLRNQLRGRTROSLPYDP-----ARDRTSLKIKWLNKC-----SG 564
Db 382 -----FYDPTTANTADSDTTL-----NLNRADAGNSTDYSG 413
Qy 565 VGVFNCOGAGMCKIEKTRIHDTG-----FGTLTAS--VCASDVDP-----ITOVAGAEML 614
Db 414 SI-VFS--GEKLSDEDAKVADNLTSLKQPVTLTACNLVLRKGVTLDTKGFTPTASSVY 470
Qy 615 GDTIVVARGSEVIRLPGKVSIVT-----LKVLEFLHFPCIOIAPISISFAIG-- 666
Db 471 MDGTTILKASTEEVTL-TGISIPVDSLGEKKVY-----IAASAASKNVALSGPIL 520
Qy 667 LIDMFNTGCAVEQVEIHNRAATKTIALSVGRGFRGVSSQRPKCVGGAETDEFDYSE 726
Db 521 LLD--NQGNAIYENHDLGKQODESFVQSLALGTAT-----TTDVPAVPTV-ATPHIGVQGT 573
Qy 727 TGLT-----TFSTVSPSEMYRMS 745
Db 574 WGMTWVDTDASTPKTKTATLAWT 596
```

RESULT 15

G86474

probable protein gypsy/Ty3 element polyprotein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86474
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: G86474

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1447 <STO>

A:Cross-references: GB:AE005172; NID:q11034946; PIDN:AA627103.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

```
Query Match 2.9%; Score 114.5; DB 2; Length 1447;
Best Local Similarity 20.2%; Pred. No. 4.8;
Matches 86; Conservative 58; Mismatches 142; Indels 139; Gaps 21;

Qy 110 NSPIITYVLLPLLEGGFAVLQGNDKNEIEICESGDNAVFED---OGLHWYMHAGTNP 166
Db 854 NAFVL---ALPYFDKQF-----MVEIDACGGGIRAVILMOKG-HP 888
Qy 167 FEVINQAVKAEKHMOTFLHREKKRLP---SCLDMFGMCTDAFYTDYTAGS---VEEG 219
Db 889 LAYISROLKGKOLHSIY---EKELLAFIFAVRKWRHYLPSHFITKTDQRSIKYLLDQR 945
Qy 220 LKSLSGSGTPRPFLIIDGMOQIENKADATECLVQ--EGAQFATRLTGIKENTKFOKKLQ 278
Db 946 LNTPYQOGMLPRLLEFDYEIDYRQGENLVADALSRYEGSEVLHMAIATIVE-CDFLKEIQ 1004
Qy 279 NNEQMSG-LKHLVHGAKOHNNVKNYVWHALAGYGVKPAATGMEHYDTALAPYQSPG 337
Db 1005 VAYESDGVLKNDIISALQOHPRAKKHYSW-----SQD 1035
```

```
Qy 338 VLGNQPDIVMDS-----LAVHGLG-----LVHPKKVFENY-----NELHAYIA 375
Db 1036 ILRRRSKIYVPADVETITKLLQWLHCSCMGGRSGRDAHQVYKSLFYKKGAVKDIQAFIR 1095
Qy 376 SCGV-----DGVKVDY-QNIITLQAGHGRV-----SLFRSY 407
Db 1096 SCGTCQCGCKSDNAAYPGILOPLPIPKIWCDSYMDPIEGLPNSGGKSVIMVVDRLSKAA 1155
Qy 408 H-----HALEA-SIASNFTDN-----GCIACMCNHTDGLYSAK-----QTAIVRASD 448
Db 1156 HVALAHFYSALTVAQAFILDNVYKHHGCPSTIVSDRDVLTFSDFWKEFFKLLQGVLELRMS 1215
Qy 449 DFYPR 453
Db 1216 AYHPQ 1220
```

Search completed: June 4, 2002, 09:06:42
Job time: 220 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 09:06:13 ; Search time 13.31 Seconds
(without alignments)
2181.793 Million cell updates/sec

Title: US-09-810-186-1
Perfect score: 3996
Sequence: 1 MTVPKISVNGKLVHGT.....TFSTPSPPEEMKRSIELQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
- score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125	3.1	982	1	NIA_NEUCR
2	117.5	2.9	4544	1	LRL1_HUMAN
3	115	2.9	928	1	PM10_CELPN
4	109.5	2.7	1338	1	VGR1_HUMAN
5	108	2.7	714	1	PBPF_BACSU
6	106	2.7	976	1	PMPE_CHLMU
7	105.5	2.6	555	1	PTB_RAT
8	103.5	2.6	587	1	PRCP_PSESR
9	103.5	2.6	2210	1	RRLP_EBOSM
10	102	2.6	2206	1	POLG_POL3L
11	102	2.6	2207	1	POLG_POL2L
12	101.5	2.5	1430	1	RPOC_XYLFA
13	101	2.5	4268	1	POLH_POL1M
14	101	2.5	4466	1	DYHC_ANTCR
15	101	2.5	4466	1	DYHC_TRIGR
16	100	2.5	2206	1	POLG_CXA21
17	99.5	2.5	815	1	RPER_HUMAN
18	99	2.5	2214	1	POLG_CXA24
19	98.5	2.5	525	1	ACU8_NEUCR
20	98.5	2.5	537	1	ACU8_NEUCR
21	98	2.5	776	1	ISOA_PSEYA
22	97.5	2.4	843	1	ARK_DROME
23	97	2.4	894	1	WPRR_BACSU
24	97	2.4	914	1	COPP_DROME
25	97	2.4	1019	1	VP3_RDVA
26	97	2.4	1536	1	GLSE_ANTSP
27	96.5	2.4	4303	1	PKD1_HUMAN
28	96	2.4	424	1	MS1L_ARATH
29	96	2.4	1403	1	BIRE_MOUSE
30	95	2.4	267	1	EBAG_FLASP
31	95	2.4	880	1	LYTD_BACSU
32	95	2.4	1019	1	VP3_RDVA
33	95	2.4	2205	1	POLG_POL2W

34	94.5	2.4	353	1	TCT2_PHRPO
35	94.5	2.4	610	1	CN42_RAT
36	94.5	2.4	1816	1	LM44_MOUSE
37	94	2.4	530	1	KPY1_RAT
38	94	2.4	551	1	CBX4_MOUSE
39	94	2.4	776	1	ISOA_PSESR
40	94	2.4	1112	1	ATB2_OREMO
41	94	2.4	2415	1	PCBM_HUMAN
42	94	2.4	4393	1	PCBM_HUMAN
43	93.5	2.3	551	1	AAT1_BACSU
44	93.5	2.3	551	1	CH62_SYNY3
45	93.5	2.3	682	1	OB66_NPVOF

ALIGNMENTS

RESULT 1					
ID	NIA_NEUCR	STANDARD:	PRT:	982 AA.	
AC	P08619;				
DT	01-AUG-1988 (Rel. 08, Created)				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	Nitrate reductase [NADPH] (EC 1.6.6.3) (NR).				
GN	NIT-3.				
OS	Neurospora crassa.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Sordariales; Sordariaceae; Neurospora.				
OX	NCBI_TaxID=5141;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=74-OR23-1A;				
RX	MEDLINE=91287699; PubMed=1829499;				
RA	Okamoto P.M., Fu Y.-H., Marzluf G.A.;				
RT	"Nit-3, the structural gene of nitrate reductase in Neurospora				
RT	crassa: nucleotide sequence and regulation of mRNA synthesis and				
RT	turnover.";				
RL	Mol. Gen. Genet. 227:213-223(1991).				
RN	[2]				
RP	PRELIMINARY PARTIAL SEQUENCE AROUND HIS-652.				
RA	Le K.H.D., Lederer F.;				
RT	"On the Presence of a heme-binding domain homologous to cytochrome b5				
RT	in Neurospora crassa assimilatory nitrate reductase.";				
RL	EMBO J. 2:1909-1914(1983).				
RN	[3]				
RP	MUTANTS.				
RC	STRAIN=74-OR23-1A;				
RX	MEDLINE=93241176; PubMed=8479443;				
RA	Okamoto P.M., Garrett R.H., Marzluf G.A.;				
RT	"Molecular characterization of conventional and new repeat-induced				
RT	mutants of nit-3, the structural gene that encodes nitrate reductase				
RL	in Neurospora crassa.";				
RN	Mol. Gen. Genet. 238:81-90(1993).				
RN	[4]				
RP	MUTAGENESIS.				
RX	MEDLINE=93360901; PubMed=8355655;				
RA	Okamoto P.M., Marzluf G.A.;				
RT	"Nitrate reductase of Neurospora crassa: the functional role of				
RT	individual amino acids in the heme domain as examined by				
RT	site-directed mutagenesis.";				
RL	Mol. Gen. Genet. 240:221-230(1993).				
CC	-1- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST				
CC	STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.				
CC	-1- CATALYTIC ACTIVITY: NADPH + nitrate = NADP(+) + nitrite + H(2)O.				
CC	-1- COFACTOR: EACH SUBUNIT CONTAINS 1 EQUIVALENT OF FAD, HEME IRON,				
CC	AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME GROUP IS				
CC	CALLED CYTOCHROME B-557.				
CC	-1- PATHWAY: NITRATE ASSIMILATORY PATHWAY.				
CC	-1- SUBUNIT: HOMODIMER.				
CC	-1- INDUCTION: ITS EXPRESSION IS HIGHLY REGULATED AND RESPONDS RAPIDLY				
CC	TO NITRATE INDUCTION AND TO NITROGEN REPRESSION.				
CC	-1- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE				

CC	N-TERMINAL DOMAIN.
CC	-1- SIMILARITY: CONTRAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC	-1- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC	C-TERMINAL DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its use
CC	by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: X61303; CAA43600.1; -.
DR	PIR: S16292; S16292.
DR	PIR: S34796; S34796.
DR	PIR: S37298; S37298.
DR	HSSP: P04166; 1B5M.
DR	InterPro: IPR001199; Cyt-B5.
DR	InterPro: IPR001834; Cyt-B5_reductase.
DR	InterPro: IPR000572; Euk-oxidored_mol.yb.
DR	InterPro: IPR001433; Oxidored_FAD.
DR	Pfam: PF00970; FAD_binding_6; 1.
DR	Pfam: PF00173; heme_1; 1.
DR	Pfam: PF00175; NAD_binding; 1.
DR	Pfam: PF00174; oxidored_mol.yb; 1.
DR	PRINTS: PR00406; CYTB5RDPASE.
DR	PRINTS: PR00363; CYTOCHROME5.
DR	PRINTS: PR00407; EMOPTERIN.
DR	PROSITE: PS00191; CYTOCHROME_B5_1; 1.
DR	PROSITE: PS50255; CYTOCHROME_B5_2; 1.
DR	PROSITE: PS00559; MOLYBDOTERIN_EUK; 1.
KW	Oxidoreductase; Flavoprotein; FAD; NADP; Heme; Molybdenum;
KW	Nitrate assimilation.
FT	METAL 240 240 MOLYBDENUM-PTERIN (POTENTIAL).
FT	METAL 295 295 MOLYBDENUM-PTERIN (POTENTIAL).
FT	DISULFID 499 499 INTERCHAIN (POTENTIAL).
FT	DOMAIN 621 691 HEME-BINDING (BY SIMILARITY).
FT	BINDING 652 652 HEME LIGAND (BY SIMILARITY).
FT	BINDING 675 675 HEME LIGAND (BY SIMILARITY).
FT	DOMAIN 718 982 FLAVIN-BINDING DOMAIN (BY SIMILARITY).
FT	NP_BIND 952 961 NADP (BY SIMILARITY).
FT	MUTAGEN 652 652 H->A: LITTLE LOSS OF ENZYME ACTIVITY.
FT	MUTAGEN 675 675 H->A: LOSS OF ENZYME ACTIVITY.
SO	SEQUENCE 982 AA; 108432 MW; B738C031B19687E CRC64;

Query Match	3.1%;	Score 125;	DB 1;	Length 982;
Best Local Similarity	22.4%;	Pred. No. 0.13;		
Matches 106;	Conservative 46;	Mismatches 178;	Indels 144;	Gaps 26

QY	194	SCLEMFQGCITWDAEYTVLTAEGVEEELKJLSOGGTFPRLLIDDMOQIEKKADA----	249
Db	496	SC-----FCMCFMD-----LEIFLSEL-----	522
QY	250	-TECLVQEGAGFATRLTGKENTKPOKKLONNOMSGIKH-----	LYVGAKOH 297
Db	523	MDESLATOPKMTIWSYLVGMNNNWEKRYVJHNSDTLREPHLPQWMLTSDGMKDYKKEGG	582
QY	298	NVKNYVYHVALAGYWGCVKPAATGMEHYDTALAYPOSPVLGNPD--IYMSLSAVHGL	355
Db	583	NLAN-----GFMGKEKVPGA-----EENVYKEEPRKESIMDEKVTRLTLEELRQHD-	629
QY	356	GLVHKKRYEN--FPNELHAVLASOGVDGVYDQNIETELGAGH--GGRVSLTRSYNHLE	412
Db	630	GEERWEFVYVNOQYVN-----GTFPLE-----GHRGGAASITGAAGDVT	668
QY	413	ASIASNFTDNGCICACMCHNTDGLYSAKQAIYVASDFTFPRPA-----SHIHSSAVYN	466
Db	669	DEFLAHSHEENKAMPMPTYHIGTLTLPSPALAKSSSTS--DPAKSDSPRLFTQSKTWN	724
QY	469	SLEFL--GEFMOPDMDMFHSL--HPADYHAAARAIGCPITYVSDKPGCNHFDLLKLTLP	524

Dn 725 SAILTFKESVSPDRIKFIFALSHPAQSI-----GLPV-----GGHIMRLPDPAPK 770

Oy 525 DGSYLRAQLP---GRPTDSL-----FVPARD-----RTSLIKTNLNC---SGYV 566
|::|| |::||
Db 771 TESTIRAVTPIPSDGTLEGLTLATVIKKIYYASPTEDINKGMOTQALDALAGAAVEFKPY 830

Oy 567 GFVNCGGAKWCI---EKKTRIHDSPTLTAAVC-ASDVDTLPVGAEMLGD 616
|::|| |::|| |::|| |::||
Db 831 GKEYVGKGVCVSNGERRKVK-----RFVMVCGSGCVTPITYVALAAYADD 876

RESULT 2
LRPL_HUMAN
ID LRPL_HUMAN STANDARD: PRT: 4544 AA.
AC Q07934;
DE 01-OCT-1996 (Rel. 34, Created)
DI 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE low-density lipoprotein receptor-related protein 1 precursor (LRP)
DE (Alpha-2-macroglobulin receptor) (AZMR) (Apolipoprotein E receptor)
DE (APOER) (CD91).
GN LRP1 OR AZMR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89210795; PubMed=3265596;
RA Herz J., Hamann U., Rognie S., Mykrebost O., Gausepohl H.,
RA Stanley K.K.;
RT "Surface location and high affinity for calcium of a 500-kd liver
RT membrane protein closely related to the LDL-receptor suggest a
RT physiological role as lipoprotein receptor.";
RL EMBO J. 7:4119-4127(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95203893; PubMed=7534747;
RA Van Leuven F., Stas L., Hilliker C., Lorent K., Umans L., Serneels L.,
RA Overbergh L., Torrekens S., Moecharts D., De Strooper B.,
RA Van den Berghe H.;
RT "Structure of the gene (LRP1) coding for the human alpha
RT 2-macroglobulin receptor lipoprotein receptor-related protein.";
RL Genomics 24:78-89(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99000832; PubMed=9782078;
RA van Leeuwen F., Stas L., Thiry E., Neijssen B., Miyake Y.;
RT "Strategy to sequence the 89 exons of the human LRP1 gene coding for
RT the lipoprotein receptor related protein: identification of one
RT expressed mutation among 48 polymorphisms.";
RL Genomics 52:138-144(1998).
RN [4]
RP PROCESSING.
RX MEDLINE=90256210; PubMed=2112085;
RA Herz J., Kowal R.C., Goldstein J.L., Brown M.S.;
RT "Proteolytic processing of the 600 kd low density lipoprotein
RT receptor-related protein (LRP) occurs in a trans-Golgi
RT compartment.";
RL EMBO J. 9:1769-1776(1990).
RN [5]
RP FUNCTION.
RX MEDLINE=91092405; PubMed=1702392;
RA Kristensen T., Moestrup S.K., Gliemann J., Bendtsen L., Sand O.,
RA Soltrup-Jensen L.;
RT "Evidence that the newly cloned low-density-lipoprotein receptor
RT related protein (LRP) is the alpha 2-macroglobulin receptor.";
RL FEBS Lett. 276:151-155(1990).
RN [6]
RP STRUCTURE BY NMR OF 1059-1100.
RX MEDLINE=9253972; PubMed=10318830;
RA Huang W., Dolmer K., Gettings P.G.W.;

FT	DOMAIN	2478	2518	EGF-LIKE 10.
FT	DOMAIN	2592	2563	LDL-RECEPTOR CLASS A 11.
FT	DOMAIN	2564	2602	LDL-RECEPTOR CLASS A 12.
FT	DOMAIN	2603	2641	LDL-RECEPTOR CLASS A 13.
FT	DOMAIN	2642	2690	LDL-RECEPTOR CLASS A 14.
FT	DOMAIN	2694	2732	LDL-RECEPTOR CLASS A 15.
FT	DOMAIN	2732	2771	LDL-RECEPTOR CLASS A 16.
FT	DOMAIN	2772	2814	LDL-RECEPTOR CLASS A 17.
FT	DOMAIN	2816	2855	LDL-RECEPTOR CLASS A 18.
FT	DOMAIN	2856	2899	LDL-RECEPTOR CLASS A 19.
FT	DOMAIN	2902	2940	LDL-RECEPTOR CLASS A 20.
FT	DOMAIN	2941	2981	EGF-LIKE 11.
FT	DOMAIN	2982	3022	EGF-LIKE 12.
FT	DOMAIN	3290	3331	EGF-LIKE 13.
FT	DOMAIN	3332	3371	LDL-RECEPTOR CLASS A 21.
FT	DOMAIN	3372	3410	LDL-RECEPTOR CLASS A 22.
FT	DOMAIN	3411	3450	LDL-RECEPTOR CLASS A 23.
FT	DOMAIN	3491	3491	LDL-RECEPTOR CLASS A 24.
FT	DOMAIN	3492	3533	LDL-RECEPTOR CLASS A 25.
FT	DOMAIN	3534	3572	LDL-RECEPTOR CLASS A 26.
FT	DOMAIN	3573	3611	LDL-RECEPTOR CLASS A 27.
FT	DOMAIN	3611	3649	LDL-RECEPTOR CLASS A 28.
FT	DOMAIN	3652	3692	LDL-RECEPTOR CLASS A 29.
FT	DOMAIN	3739	3778	LDL-RECEPTOR CLASS A 30.
FT	DOMAIN	3781	3823	EGF-LIKE 14.
FT	DOMAIN	3824	3861	EGF-LIKE 15.
FT	DOMAIN	4147	4183	EGF-LIKE 16.
FT	DOMAIN	4196	4232	EGF-LIKE 17.
FT	DOMAIN	4232	4268	EGF-LIKE 18.
FT	DOMAIN	4268	4304	EGF-LIKE 19.
FT	DOMAIN	4304	4340	EGF-LIKE 20.
FT	DOMAIN	4340	4375	EGF-LIKE 21.
FT	DOMAIN	4373	4409	EGF-LIKE 22.
FT	SITE	3940	3943	PROTEOLYTIC SITE FOR PROTEOLYTICAL PROCESSING (POTENTIAL).
FT	SITE	4502	4507	ENDOCYTOSIS SIGNAL (POTENTIAL).
FT	DISULFID	27	40	BY SIMILARITY.
FT	DISULFID	34	53	BY SIMILARITY.
FT	DISULFID	47	64	BY SIMILARITY.
FT	DISULFID	72	85	BY SIMILARITY.
FT	DISULFID	79	98	BY SIMILARITY.
FT	DISULFID	92	108	BY SIMILARITY.
FT	DISULFID	115	124	BY SIMILARITY.
FT	DISULFID	120	133	BY SIMILARITY.
FT	DISULFID	135	148	BY SIMILARITY.
FT	DISULFID	154	164	BY SIMILARITY.
FT	DISULFID	160	173	BY SIMILARITY.
FT	DISULFID	175	188	BY SIMILARITY.
FT	DISULFID	478	493	BY SIMILARITY.
FT	DISULFID	489	504	BY SIMILARITY.
FT	DISULFID	506	519	BY SIMILARITY.
FT	DISULFID	807	818	BY SIMILARITY.
FT	DISULFID	814	827	BY SIMILARITY.
FT	DISULFID	829	842	BY SIMILARITY.
FT	DISULFID	854	866	BY SIMILARITY.
FT	DISULFID	861	879	BY SIMILARITY.
FT	DISULFID	873	890	BY SIMILARITY.
FT	DISULFID	895	907	BY SIMILARITY.
FT	DISULFID	902	920	BY SIMILARITY.
FT	DISULFID	914	931	BY SIMILARITY.
FT	DISULFID	936	948	BY SIMILARITY.
FT	DISULFID	943	961	BY SIMILARITY.
FT	DISULFID	955	971	BY SIMILARITY.
FT	DISULFID	976	989	BY SIMILARITY.
FT	DISULFID	984	1002	BY SIMILARITY.
FT	DISULFID	996	1011	BY SIMILARITY.

Query Match

Best local Similarity 2.9%; Score 117.5; DB 1; Length 4544;

Matches 98; Conservative 51; Mismatches 178; Indels 145; Gaps 25;

71 CPEPKKMKMMPQBMST-----CPVPLFMQMTTFRGGRCTMMDDNDIVYVITR 101

```

Db 3628 CLR-----GTCSQLCNNTKGHLCCARNFM---KTHNCKAESE--YQYLTA 3673
QY 122 LEGOFRAVLQGNDRKNEITEICLSEGDNAVETD-QGLHM-----VY---MHAGTNPFEVINA 173
Db 3874 DQNEIRSLFPHPHSAVEQAFO-GDESVRIDAMVHKAGRWVTNMHTGTISRSRPPA 3932
QY 174 VKAV--EKHMO-----TFLHREKKRLPSCIDMFQGTWDAFYTDVAEGVE-----EG 219
Db 3933 APPTSNHRRROIDRGVTHLNIISGLKMPRGIA-IDWVAGVNYWTDSDRDVYEAQMKGEN 3991
QY 220 LKLSLQSG-PPRRLIID-----DGM---OOIENKAKDAR--ECLVGAQOFATRLT 265
Db 3992 RKTLSGIDEPHAIIVDPLRGITWSDMGNHFKIETAAADGTIRETLVDNIDNPTGL- 4050
QY 266 GIKENTKFOKKLONNEQMSGKHLVHGAKOHNVKKNYVYVHVALAGVGVKPAATGMEHY 325
Db 4051 -----AVDYHNERL-----YWADAKLSVIG--- 4070
QY 326 DTAALAVPQSGVIGNOPDIYMDSLAVHGLGVHPKVFNFYVNLHAYLASCGVGVKVD 385
Db 4071 -----SIRLNGTDEIVADSKR---GLSHPFSDVFEEDYI-----GVTYI 4108
QY 386 VONITETLGAHGGRVSLTRSYHHALEASI-----ASNFTDNG-----CI-----A 426
Db 4109 NNRVFKIKRFGHSPVYVNLTGSLSHASDVLYLHOKRQPEVTNPDCKRKCCEMCLLSPSGPV 4168
QY 427 CMCNHTDGLYSAKOTAIIVRASDDFPRDPASHTTHISSAVNSLFLGEMQP 478
Db 4169 CTCPRGKRLDNG--TCVPVPSPTPPDPAPRPGTNCLOCFNGSCFLNAROP 4218

RESULT 3
PM10_CHLPN STANDARD; PRF: 928 AA.
ID PM10_CHLPN
AC Q9RB65; Q9RB64; Q9S6P2; 086163;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpl0 precursor (Polymorphic membrane
DE protein 10) (Outer membrane protein 5).
GN PM10 OR OMP5 OR CP0303.
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RX MEDLINE=20007584; PubMed=10539856;
RA Christiansen G., Boesen T., Hjerno K., Dagaard L., Mygind P.,
RA Madsen A.S., Knudsen K., Falk E., Birkelund S.;
RT "Molecular biology of Chlamydia pneumoniae surface proteins and their
RT role in immunopathogenicity."
RL Am. Heart J. 138:S491-S495(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linier K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hatakeyama H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

```

```

RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE OF 1-914 FROM N.A.
RC STRAIN=CWL-029/VR-1310;
RX MEDLINE=99081766; PubMed=9864239;
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding 97- to 99-kilodalton outer
RT membrane proteins of Chlamydia pneumoniae."
RL Infect. Immun. 67:375-383(1999).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC CC
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC CC
CC EMBL: AJ133034; CAB37071.1; -
CC DR EMBL: AE002192; AAF38160.1; -
CC DR EMBL: AF002546; BAA98657.1; -
CC DR EMBL: AJ001311; CAA04671.1; -
CC DR PHC1-2DPAGE: 086163; -
CC DR TIGR: CP0303; -
CC DR InterPro: IPR003357; OMP.
CC DR Pfam: PF02385; OMP; 1.
CC KW Outer membrane; Signal; Multigene family.
CC FT SIGNAL 1 25 POTENTIAL.
CC FT CHAIN 26 928 PROBABLE OUTER MEMBRANE PROTEIN PMP10.
CC SQ SEQUENCE 928 AA; 97229 MW; 0590D5206A1DD0E1 CRC64;

Query Match 2.9%; Score 115; DB 1; Length 928;
Best local similarity 24.2%; Pred. No. 0.72;
Matches 107; Conservative 39; Mismatches 145; Indels 152; Gaps 23;

QY 338 VLGNOPDIYVMDSLAVHGLGVHPKVFNFYVNLHAYLASCGVGVKVDVONITETLGAAGH 397
Db 271 ISGQSVTFESGNOAVANGAIVAKL-----TLASGGGGISFS--NNIVGGTTAGN 320
QY 398 GGRVSLTRSYHHALEASISNFTDNGCIACMCNHTDGLYSAKOTAIIVRASDDFPRDPAS 457
Db 321 GGAISILAAEGCSLSAE--AGDITFNG-----NALVATF---PQTRK 358
QY 458 HTTHISSAVNSLFLGEMQPPMDMFRSLHPADYHAAARAIGGCPYVSGDKGNHNPDL 517
Db 359 NSIDIGSTAKIT-----NLRAISGHSIF----- 381
QY 518 LKKIVLPDGSVLRAQLPGRPTDLSFVDP-----ARDRTSLKIVMLNKC-----SG 564
Db 382 -----FYDPTTANTADSDTL---NLNKADAGNSTDYG 413
QY 565 VVGVFNCQAGWCKIEKTRIHDT-----EGTLTAS--VCASDVLD---ITOVAGAEML 614
Db 414 SI-VFS--GKLSDEDAKVADNLTSLKCPVTLTAGMLVLRGVTLDTKGFTOTAGSVYI 470
QY 615 GDTIVVYRSGEYVRLPKGVISPYT-----LKVLEFELFHFCITQELAPISFAAG--- 666
Db 471 MDAQTLKASTEEVTL--TGHSIPVDSIGEGKKV-----TASASKNVALSGPIL 520
QY 667 LLDENFGAVEQVEIHNRATKTIALSVGRGRFGVSSQRPDLKCVGGAETDFNYDSE 726
Db 521 LLD--NQGNAVENHDLKTDQDFVQLSALGTAT-----TTDVAVPPTV--ATPHVGVGGT 573
QY 727 TGLT-----TFISVPSPEMTRMS 745
Db 574 WGMTWVDVDTASTPKTKTATLAWT 596

```



```

QY 206 APTDTVTAEGVEGLKSLSGGTP-PRFLIIDGMOQIENKA-----KQATECL 253
D 346 -----TVAGRSRYRLSMKAKKAPPEPVWLKDKDLPAATEKSATYLRGYSLIKDYTE-- 397
QY 254 VQGAQAFATLTGKIKENTKFKQKLONNEMQSLKHLVHGAKOHNYKNYVWHALAGYNG 313
D 398 --EDAGNYTLLSLIKOSNVE-----KNLT--ATLVNKKPQIYEKAVSSF-- 438
QY 314 GVPRA--ATGMEHYDALAPVPSGV-----LGNOPIVIMSLAV 352
D 439 -PPDALPLPSRQILCTCTAGIPQPTIKWPHPCNNHSEARDFCSNNESEFILLADSN 497
QY 353 HG-----LGLVHKKKAPNYNEIHLV--ASGVGVSKVDY-QNIIEITLQAG----- 396
D 498 MGNRIESTORMAIIIEK-----NKMASLVVADSRISGIYCIASMKVYGKRNISFY 551
QY 397 -----HGRVSL-----TRSYHHA----- 411
D 552 ITDVPNGFHNLEKMPTEGEDKILSCVYNKFLRYDTWILLRTVNNKNTMHSISKQKMAI 611
QY 412 --EASIASNFT-----DNGCIACMCHNT--DGLYSAKQTAIVRASDDFYPRDASHT 459
D 612 TKHSITLINTLIMVSLQDSGYACRARNYTGEEILQKKEITIRQEAPEYLLRNLSDT 671
QY 460 IHISVAVNSLFLGEFQOPMDMFHSLHPADYHAAARAIIGGCPYVSDKRGHNFDLTK 519
D 672 VAISSTTLDCHANGVPEPQITWFKNNK-----IQDEPG----- 706
QY 520 KLIVPDGSLVLAQPLGRPTRLSEVPDARDTSLKIMNLKSGVGVNCGAGWKCI 579
D 707 -IIIGPS-----STLFIEVTEDE-----GVYHCKAT----- 734
QY 580 EKKTRIHDTSPGTLTASVASCADVLIT 606
D 735 NQKGSVESSAYLTVOGTSDKSNLELT 761

RESULT 5
PBPF_BACSU STANDARD; PRT; 714 AA.
AC P38050;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Penicillin-binding protein 1F (PBP-1F).
GN PBP OR PONA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
CX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98240224; PubMed=9579061;
RA Nockack M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,
  Wedder H., Venema G., Bron S.;
  "The 172 kb prkA-addAB region from 83 degrees to 97 degrees of the
  Bacillus subtilis chromosome contains several dysfunctional genes,
  the glyB marker, many genes encoding transporter proteins, and the
  ubiquitous hit gene.";
  Microbiology 144:859-875(1998).
RN [2]
RP SEQUENCE OF 1-129 FROM N.A.
RC STRAIN=168;
RX MEDLINE=93328693; PubMed=8335642;
RA Popham D., Setlow P.;
  "Cloning, nucleotide sequence, and regulation of the Bacillus
  subtilis pbpF gene, which codes for a putative class A
  high-molecular-weight penicillin-binding protein.";
  J. Bacteriol. 175:4870-4876(1993).
RN [3]
RP SEQUENCE OF 122-714 FROM N.A.
RX MEDLINE=93094140; PubMed=1459957;

```

```

RA Hansson M., Hederstedt L.;
RT "Cloning and characterization of the Bacillus subtilis hemyH gene
  cluster, which encodes proteoheme IX biosynthetic enzymes.";
RL J. Bacteriol. 174:8081-8093(1992).
CC -1- FUNCTION: CELL WALL FORMATION. MAY BE INVOLVED IN OUTGROWTH OF THE
  GERMINATED SPORE OR IT COULD FUNCTION IN THE SYNTHESIS OF THE GERM
  CELL WALL.
CC -1- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION REMAINS CONSTANT DURING VEGETATIVE
  GROWTH, DECREASES DURING EARLY SPOULATION, AND IS INDUCED IN THE
  FORESPORE DURING LATE SPOULATION.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
  TRANSGLYCOSYLASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
  TRANSEPTIDASE FAMILY.
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb.ch/announce/
  or send an email to license@isb.slb.ch).
CC -----
DR EMBL: Y14083; CAA74517.1; -.
DR EMBL: L10630; AAA71942.1; -.
DR EMBL: M97208; AAA22516.1; -.
DR EMBL: Z99109; CAB12851.1; -.
DR PIR: A40614; A40614.
DR Subtilist; BG10428; pbpF.
DR InterPro: IPR001264; Transglycosyl.
DR InterPro: IPR001460; Transpeptid.
DR Pfam: PF00912; Transglycosyl. 1.
DR Pfam: PF00905; Transpeptidase. 1.
DR Prodom: PD001895; Transglycosyl. 1.
DR Peptidoglycan synthesis; Cell wall; Transmembrane; Signal-anchor;
  Complete proteome.
KW DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
  TRANSMEM 13 33 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
  POTENTIAL.
FT DOMAIN 34 714 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 49 217 TRANSGLYCOSYLASE.
FT DOMAIN 297 592 TRANSEPTIDASE.
FT ACT SITE 359 359 ACYLATED BY PENICILLIN (BY SIMILARITY).
SQ SEQUENCE 714 AA; 79278 MW; 08D96718C84B434 CRC64;

Query Match 2.7%; Score 108; DB 1; Length 714;
Best local similarity 19.6%; Pred. No. 1.7;
Matches 121; Conservative 77; Mismatches 239; Indels 180; Gaps 26;

QY 156 HWY-MHAGNPREVINQAVK-----AVEKHQTF--LH--REKKRLPSCLDW 198
D 172 HGAYGIQAAASHY--FNKEKGLTVSEGANLAIIPKAPSTYSPILHPDKNKERRRDTIICM 229
QY 199 FGMCWDATFYDTVTAEGVEGLKSLSGGTPPRFLIIDGMOQIENKARDATECIVQEGA 258
D 230 MNOGATISAKENATYAGCRTGLGHVKKQSETPWDSYIDLVIKRAEDKYSISGROLQGGY 289
QY 259 QFATRLTGKIKENTKFKQKLONNEMQSLKHLVHGAKOHNYKNYVWHALAGWGVKPA 318
D 290 TIKVPLDSKLOKTAYQ-VMEKGSYYPGTDONAGSAVFINKTK-----GGVEEA 337
QY 319 ATGMEH-----YDPAALAPVPSGVILGNOPDIVMDSLAVHGLGVHKKYFNFEIHLAVL 374
D 338 IGGDDYTSKGYNRVAV-----KQPSGTIKPLAVIGPAN--QEKFKFYSILKDEL 386
QY 375 ASGVGVSKVDYQNIIEITLQAGHGGRVSLRSYHHALEASIASNFTDNGCIACMCHNTG 434
D 387 QSYG-----DYTPKNYSRYEGEVTMSDALTYGKNAP-----AWTLINEIG 427

```


[illegible]

[illegible]

FT ACT_SITE 385 385 PROBABLE.
 FT ACT_SITE 502 502 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 543 543 PROBABLE.
 FT DISULFID 352 391
 SO SEQUENCE 587 AA; 61072 MW; E193DB82C225829A CRC64;

Query Match 2.6%; Score 103.5; DB 1; Length 587;
 Best Local Similarity 21.8%; Pred. No. 2.9; Mismatches 204; Indels 159; Gaps 26;
 Matches 114; Conservative 47;

QY 243 ENKAKATECLVOEG-AQEPATRLTGIKETKFOKKLONNEOMSGELKHLVGAQH-----296
 DB 82 EAOLKOLAQAVNPGNAQEG-----KFLKRRQFLSOFAPTAQOAVAHLRKNG 131
 QY 297 ----HWKKNVYVHVALGAWGVKPA-ATGMEHYD-----ALAVPVSQVGLNQDPI 345
 DB 132 FVAIHVYVPRNLISA-DGSAQAVKAFNTPVYQLNGKAGYANPAAPVODLG---EI 187
 QY 346 VMDSLAVHGLVHPKRVENFYNELHAYLASGVDSKVQVONIIETLGAGHGRVSLFR 405
 DB 188 VGSVLGLQVNTTRAHP-----LKGESSAKTLAAG-----TA 220
 QY 406 SYHHALEASTASFTDNGCLACKCHNTDGLYSAKQTAIVRASDDFYPDPASHTIHSSV 465
 DB 221 KGNHPTPEPT-----IYDASSAPTA-----AMTVGIIIT 250
 QY 466 AYSLFLGEMOPDMDFSLHPADYHAAARAIGGC-PIYVSDKRGNNHPLKLVLP 524
 DB 251 GGVSQLQLOLQ-----FTSANGLASVNTQTQTGSSNGSDYDDOGGOWLDQSOTV- 304
 QY 525 DGSVLRAQLPGRPTDLSLV---DPAEDRTSLKTIWNLKSGVGVENCQAGCMCKIEK 581
 DB 305 -SSA-----GGAVOQLLFYMAQDQASGNTGLQAFNQAVSNVAKVINV-SLGCCEADA 356
 QY 582 KRRIHDSGTLTASVCASDVLLITQVGAEM-----LGDITIV-----AYRSEVITRL 630
 DB 357 -----NADGTLQAE-----DRIFATAAQQOTFSVSSGDEGVYECUNRGYDGSY-- 402
 QY 631 PKGSIPTVLTLEFELHFCPILOELAPISPAIIGLDMFT--GGAVEQVEIHNRAATK 689
 DB 403 --SVSNP-----ASSPNVIAVGGTTLTTSAGAASNETWNEGLDS 441
 QY 690 TIALSVRGGRGVYSSQRLKCVVGAEE---TDVNDSETG 728
 DB 442 NCKLWATG-GGYSVESKPSQSVSGTPGRRLPDISPDAAG 484

RESULT 9
 RPL_EBOSM STANDARD: PRT; 2210 AA.
 ID RPL_EBOSM 066802;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein) (L protein).
 GN L.
 OS Ebola virus (strain Sudan Maleo-79) (Ebo).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
 CC Filovirus.
 NCBI_TaxID=128949;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sanchez A., Trapier S., Nichol S.T.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACTIVE POLYMERASE. IT
 CC MAY FUNCTION IN RNA SYNTHESIS, CAPING, AS WELL AS METHYLATION OF
 CC CAPS, AND POLY(A) SYNTHESIS. ALSO INVOLVED IN TRANSCRIPTIONAL RNA
 CC EDITING.
 CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -I- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS

CC OF RHABDOVIRUSES AND PARAMYXOVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U23458; AAA79970.1;
 DR InterPro: IPR001016; Paramyx_RNA_pol.
 DR Pfam: PF00946; Paramyx_RNA_pol.1.
 KW transferase, RNA-directed RNA polymerase.
 SO SEQUENCE 2210 AA; 251255 MW; F50E6B624951AE44 CRC64;

Query Match 2.6%; Score 103.5; DB 1; Length 2210;
 Best Local Similarity 22.2%; Pred. No. 20; Mismatches 195; Indels 123; Gaps 25;
 Matches 107; Conservative 56;

QY 315 VKPATGMEH--YDTALAVPQSPGVGLNQDPIYMDSLAVHGLVHP---KKVFNFYN 368
 DB 483 IKDRATAVEQTCMDA-----VFEPNVLGSPRYFNFKRVPEQFLQEDSIESVLYQAQ 537
 QY 369 ELHATVASCQVQKAVYVONIIETLGAGHGRVSLTSSYHHALEA---SIASFTDNGC 424
 DB 538 ELRYLLPQNRNFSFLKEKEL--NVGRTEGKLPYLITRVVOTICBALLADGLAKAFPSNM 595
 QY 425 IACMCHNTDGLYSAKQTAIVRASDDFYPDPASH-TIHSSVA-----YNSLFLGEMOP 478
 DB 596 VVTEBECKESL-LHQASMHHTSDP-----GEATYRGSSFEVDLEKYNLAERFETAP 648
 QY 479 -----DMDMFHSLHPADYHAAARAIGCPIYVSD--KPSNHFEDLLKTL 521
 DB 649 FIKYCNCQCYGVANFEDW--MHFLIPQ-----CVMHNSDYVNP-PHVNTLENRE 693
 QY 522 VLPRG-SVLRQQLRG-KPTBDSLEVPDARDTSILKT-----WNLNKSGVGVGF 569
 DB 694 YPPGSPVSRHGLIGLQKMTSISCAQISLVEIKTFKRLSAVAGNOCITVLSVF 753
 QY 570 NCGAGCMCKIEKTRIHDSPTGLTASVCASDVLLITQVGAEMLGD-TIYVYRSGEVI 628
 DB 754 PLESSP-----NEQRCADENARVAASLAK-----VISACGIPLKPDETVH---SGFIT 801
 QY 629 RLKP-----GVSIPTLVY-----LEFELP-----HFCPIQET 656
 DB 802 FGPKQYINGIDLPQSLKTAARMAPLSDAIFPDLOGTLASIGTAFERSISFTRHILPSRVA 861
 QY 657 APSISFAIGLDM-----FMTGGAVEQVEIHNRAATITLSVRGRGRFVYSSQRLK 712
 DB 862 AAFHTTFSVRLLHGHGIFHGSGLQALINKPLDFGTALSLAVPVLGSLFLNPERK 921
 QY 713 V 713
 DB 922 L 922

RESULT 10
 POLG_POL3L STANDARD: PRT; 2206 AA.
 ID POLG_POL3L 084783; 084784; 084785; 084786; 084787; 084788; 084789;
 AC 084790; 098592; 098593; 098594;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins
 DE P2A TO P2C, P3A; Genome-linked protein VEG; Picornain 3C
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 DE (EC 2.7.7.48)].
 OS Poliovirus type 3 (strains P3/Leon/37 and P3/Leon 12A11[B]).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 CC Enterovirus.

OX NCBI_TaxID=12088;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3/LEON/37;
 RX MEDLINE=84170338; PubMed=6324200;
 RA Stanway G., Hughes P.J., Mountford R.C., Reeve P., Minor P.D.,
 Schild G.C., Almond J.W.;
 RT "Comparison of the complete nucleotide sequences of the genomes of
 the neurovirulent poliovirus P3/Leon/37 and its attenuated Sabin
 vaccine derivative P3/Leon 12a1b.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:1539-1543(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3/LEON 12A11B;
 RX MEDLINE=83299239; PubMed=6310508;
 RA Stanway G., Cann A.J., Hauptmann R., Hughes P.J., Clarke L.D.,
 Mountford R.C., Minor P.D., Schild G.C., Almond J.W.;
 RT "The nucleotide sequence of poliovirus type 3 Leon 12 alb: comparison
 with poliovirus type 1.";
 RL Nucleic Acids Res. 11:5629-5643(1983).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-878.
 RX MEDLINE=95120467; PubMed=7820548;
 RA Grant R.A., Hirmath C.N., Filman D.J., Syed R., Andries K.,
 Hogle J.M.;
 RT "Structures of poliovirus complexes with anti-viral drugs:
 implications for viral stability and drug design.";
 RL Curr. Biol. 4:784-797(1994).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1-878.
 RA Syed R., Filman D.J., Hogle J.M.;
 RT Submitted (MAR-1995) to the PDB data bank.
 RL
 CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 Q/Q SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN SABIN VACCINE P3/LEON/37 IS
 CC SHOWN.
 CC -1- MISCELLANEOUS: THE STRAIN SABIN VACCINE P3/LEON/37 IS THE
 CC PROGENITOR OF THE STRAIN SABIN VACCINE P3/LEON 12A11B.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL: K01392; AAA46914.1; -
 DR EMBL: X00925; CAA25444.1; -
 DR PIR: A03900; GNNYAP.
 DR PDB: 1PIV; 20-JUL-95.
 DR PDB: 1PVC; 15-SEP-95.
 DR PDB: 1YBA; 11-JUL-96.
 DR PDB: 1YBB; 11-JUL-96.
 DR PDB: 1YBC; 11-JUL-96.
 DR PDB: 1YBE; 11-JUL-96.
 DR MEROPS: C03.001; -
 DR InterPro: IPR001199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rhy.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.

DR Pfam: PF00073; Rhy; 3.
 DR Pfam: PF00680; RNA-dep. RNA-pol; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KW Polypeptide; Coat protein; Core protein; Transferase;
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
 KW 3D-structure.
 FT CHAIN 2 69 COAT PROTEIN VP4.
 FT CHAIN 70 340 COAT PROTEIN VP2.
 FT CHAIN 341 578 COAT PROTEIN VP3.
 FT CHAIN 579 878 COAT PROTEIN VP1.
 FT CHAIN 879 1027 CORE PROTEIN P2A.
 FT CHAIN 1028 1124 CORE PROTEIN P2B.
 FT CHAIN 1125 1453 CORE PROTEIN P2C.
 FT CHAIN 1454 1540 CORE PROTEIN P3A.
 FT CHAIN 1541 1562 GENOME-LINKED PROTEIN VP6.
 FT CHAIN 1563 1745 PICORNAIN 3C.
 FT CHAIN 1746 2206 RNA-DIRECTED RNA POLYMERASE P3D.
 FT LIPID 2 MYRISTATE.
 FT ACT_SITE 1709 1709 PROTEASE (POTENTIAL).
 FT ACT_SITE 1723 1723 PROTEASE (POTENTIAL).
 FT VARIANT 431 431 S -> F (IN P3/LEON 12A11B).
 FT VARIANT 864 864 K -> R (IN P3/LEON 12A11B).
 FT VARIANT 908 908 T -> A (IN P3/LEON 12A11B).
 SO SEQUENCE 2206 AA; 246163 MW; 4766B15C861F66D3 CRC64;

 Query Match 2.6%; Score 102; DB 1; Length 2206;
 Best Local Similarity 21.0%; Pred. No. 27;
 Matches 122; Conservative 67; Mismatches 207; Indels 184; Gaps 33;

 QY 10 NDGKLVVH-GKTLITGVPPNVVLTG-----SGR-----GLY 40
 DB 1660 NDGVLIVNSKRYNMVPPVGAIVTEQCYLIMGQFARILMYNPPRAGCGGVTCTGKV 1719
 QY 41 TGAEVGATASHSKSLHVPFGVLEGLREKCCERFKMMWTFORMGTCGRVPLETFQMLIE 100
 DB 1720 IGHWVGNGSH-----GFAAAL-----KRSYFTOSG-----ETQMMR-P 1753
 QY 101 SKSEFDGENSPILITVLPLE-GQFRAVLQDNDENEIEICSGDNAVETDGLHWY 159
 DB 1754 SKRA-----GYPLINAPTKLPSAFHYFEVKEPAV-----LTKNDPLKIDFEALIS 1805
 QY 160 MHAGTNEFEYINQAVAKVEKMQTFLEKRRPLRSLDMFGWCTWDAFY-TDVTAEGVEE 218
 DB 1806 KYGKNKTEVDEYKMEKVDHYAGQLMSLDISTBOMCLE-----DAMYGTD----- 1850
 QY 219 GLKSL--SCGGTPPRFLIIDGWOQLEK---ADADTECLVQEGAOAFATRLTGIKENT 271
 DB 1851 GLEALDLSSTAGYR--YVAMGKKRRDILNKOTRDTREKORLLDAYGINLP-VTVYKDEL 1907
 QY 272 KPKOKLQNNQMSGKLVKLVGAKOHNVKVVYVHVALAGVGVKPAATGMEHYDTALAY 331
 DB 1908 RSKTKVE-----QKSLTIASSLNSV-----AMBAFGLVYAF-----H----- 1944
 QY 332 PVQSPGV-----LGNQDPIYMSLAVHGLGLVHPKKVFNF-YNELHAYIASCGVDGVKVD 385
 DB 1945 --RNPQVVTGSAGVCDPDLFWSKIPV-----LMEEKLFADYGYDASLSPAMFEALIKVY 1997
 QY 386 VQNTIEFLGAGHGRVS-----LRSYIHLEAASIASFTDNGCIACMCHNTDGLYSAGKT 441
 DB 1998 LEKI-----GFGDRVDYIDLHNS-HILYKNKI--YCVKGGMPSCGSGSTISFNSMINN 2047
 QY 442 AIYVA-----SDPE-----YPRD-----PASHTH 461
 DB 2048 LITITLLKITYKGIIDDLKMAVAGDVYASVYHEVDASLSAGSKDYGLTWPAPKASAT 2107
 QY 462 ISSVAY-NSLFLGFMQPDWD---MFHSLHPAADYHAAAR 497
 DB 2108 FEVTVMENVTFLKRFRADEKYPFLIHPVMPKMEIHSIR 2147

RESULT 11
POLG_POL2L STANDARD; PRT: 2207 AA.
AC P06210:
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE Genome polypeptide (Contains: Coat proteins VP1 no VP4; Core proteins P2A to P2C, P3A; Genome-linked protein VP3; Picornain 3C (EC 3.4.22.28) (Protease 3C); RNA-directed RNA polymerase p3D (EC 2.7.7.48)).
DE Poliovirus type 2 (strain Lansing).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.
OX NCBI_TaxID=12084;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=66115399; PubMed=3003384;
"la Monica N., Merlam C., Racanelli V.R.;
Mapping of sequences required for mouse neurovirulence of poliovirus type 2 Lansing."
RT J. Virol. 57:515-525(1986).
RL
CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN O/C SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
CC -1- PM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb.sib.ch).

CC EMBL: M12197; AAA46912.1; .
DR PIR: A29507; GNNY5P.
DR HSSP: P03299; IPOV.
DR MEROPS: C03.001; .
DR MEROPS: C03.020; .
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhv.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; Rhv; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
KW Polypeptide; Coat protein; Core protein; Transferase; RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
FT CHAIN 2 69
FT CHAIN 70 340
FT CHAIN 341 578
FT CHAIN 579 879
FT CHAIN 880 1028
FT CHAIN 1029 1125
FT CHAIN 1126 1454
FT CHAIN 1455 1541
FT CHAIN 1542 1563
FT CHAIN 1564 1746
FT CHAIN 1747 2207
RNA-DIRECTED RNA POLYMERASE 3D.

FT LIPID 2 2 MYRISTATE.
FT ACT_SITE 1710 1710 PROTEASE (POTENTIAL).
FT ACT_SITE 1724 1724 PROTEASE (POTENTIAL).
SQ SEQUENCE 2207 AA; 245829 MM; 281E2070B7D44F99 CRC64;

Query Match 2.6%; Score 102; DB 1; Length 2207;
Best Local Similarity 20.7%; Pred. No. 27;
Matches 119; Conservative 68; Mismatches 214; Indels 174; Gaps 30;

QY 10 NDGKLVNH-SKTIITGVDMVNLTPG-----SGR-----GIV 40
DB 1661 NDGVLINTSKYPMYVPGAVTEGGLNSGQTARTIMYPTTAGCCGVTCTGAV 1720
QY 41 TGAFTVATASHSKSLHVEPVGVLGLRFMCFFRKIMMTQRMGTGRDVPLETFMLE 100
DB 1721 IGMHVGNGSH-----GFAAL-----KSYFTQSOG-----EIQMR-P 1754
QY 101 SKSEIDGENSPITTYVLLPGLSGPRAVLQGDKNIEICLSEGNNAVETDGLMVTM 160
DB 1755 SKEVGYVINAP-SKTKLEP---SAHYVEGVKEPAV--LTKSDPRUKTDEEAFISK 1807
QY 161 HAGTNPETVINAQKAVKEMQTFLEHREKKRLPSCLDMFGWCTWDAFY-TDVTAEVGE 219
DB 1808 YGKMKITEVDEYKKAVDHAGQLMSLDINTEQMCLE-----DAMYGTD-----G 1852
QY 220 LKSL-----SOGGTPPRFLIIDGMOQIEBK---AKDTECLVOEGAQFATRLTGIRENTR 272
DB 1853 LEALDISTSAGVP--YVAMGKKRRDILNKOTRDTKEMQRLDLYGINPL-VTVYKDELR 1909
QY 273 FOKRLQNNEMSLKHLHYHAKQHNHVKVYVHNAIAGVWGKVPRAQMTMHTDRLAIVP 332
DB 1910 SKTKVE-----QCKSRILEASLINDSV-----ARMAFGNLYAAE---H----- 1945
QY 333 VQSPGV-----IGNPDIVWDSIAVHGLGVHKKVFN-YNELHAYIASCGVDGKVVDV 386
DB 1946 -KMPGVVTSAGVCCDPLDFMSKIPV-----LMEEKLFADYIGYDLSLSPAFEAIAKWL 1999
QY 387 QNIETLIGAGHGRVSLTISYHHALEASIASNFTDNGCIACMCNTDGLYSAKQAIKVA 446
DB 2000 EKI-----GFGDRVDYIDYLNHSHLYKNKTYCVAGMPSGCSGTSIPNSMINILIRT 2053
QY 447 -----SDDF-----YPRD-----PASTHIHSSVA 466
DB 2054 LLKTKTGIDDLKMIANGDVYASYPHEVDASLSAOSGKGYGLTMPADKSAFETV 2113
QY 467 Y-NSLFLGERMOPDWD---MFHSLRPADYHAAAR 497
DB 2114 WENVTEFLKRRFRADEKYPFLVHPVMPMKETIHESIR 2148

RESULT 12
RPOC_XYLFA STANDARD; PRT: 1430 AA.
ID RPOC_XYLFA
AC Q9PA87;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR XF2632.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
RN NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M., Alvarenga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,

CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- CAUTION: SEE ALSO VERSION 1 OF THIS PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: V01149; CAA24461.1; ALT_SEQ.
DR EMBL: V01149; CAA24462.1; ALT_SEQ.
DR EMBL: V01149; CAA24463.1; ALT_SEQ.
DR EMBL: V01149; CAA24464.1; ALT_SEQ.
DR PIR: A03898; GNNY2P.
DR PDB: 2PLV; 15-JUL-93.
DR PDB: 1FPT; 31-MAR-95.
DR PDB: 1POV; 07-DEC-95.
DR PDB: 1VBD; 11-JUL-96.
DR PDB: 1AL2; 19-NOV-97.
DR PDB: 1AR6; 03-DEC-97.
DR PDB: 1AR7; 03-DEC-97.
DR PDB: 1AR8; 03-DEC-97.
DR PDB: 1AR9; 03-DEC-97.
DR PDB: 1ASJ; 03-DEC-97.
DR PDB: 1PO1; 03-DEC-97.
DR PDB: 1PO2; 03-DEC-97.
DR MEROPS: C03.020; -.
DR -MEROPS: C03.020; -.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhv. 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; rhv. 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
KW Polypeptide; Coat protein; Core protein; Transferrase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
KW 3D-structure.
FT INT MET 0 0
FT CHAIN 1 68 COAT PROTEIN VP4.
FT CHAIN 69 340 COAT PROTEIN VP2.
FT CHAIN 341 578 COAT PROTEIN VP3.
FT CHAIN 579 880 COAT PROTEIN VP1.
FT CHAIN 881 1029 CORE PROTEIN P2A.
FT CHAIN 1030 1126 CORE PROTEIN P2A.
FT CHAIN 1127 1455 CORE PROTEIN P2C.
FT CHAIN 1456 1542 CORE PROTEIN P2C.
FT CHAIN 1543 1564 GENOME-LINKED PROTEIN VP6.
FT CHAIN 1565 1746 PICOCHAIN 3C.
FT CHAIN 1747 2208 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 1 1 MYRISTATE.
FT ACT_SITE 1711 1711 MYRISTATE.
FT ACT_SITE 1725 1725 PROTEASE (POTENTIAL).
FT STRAND 3 6
FT STRAND 25 28
FT HELIX 35 37
FT HELIX 45 45
FT STRAND 49 49
FT TURN 49 49

FT STRAND 50 53
FT HELIX 56 56
FT TURN 62 63
FT TURN 74 75
FT TURN 79 80
FT STRAND 82 86
FT TURN 87 88
FT STRAND 89 93
FT STRAND 96 101
FT HELIX 102 104
FT TURN 112 114
FT STRAND 122 122
FT HELIX 125 125
FT TURN 128 128
FT STRAND 132 133
FT STRAND 137 140
FT TURN 141 142
FT STRAND 146 150
FT TURN 151 151
FT HELIX 152 154
FT TURN 155 156
FT HELIX 158 166
FT STRAND 167 179
FT TURN 184 185
FT STRAND 186 196
FT TURN 197 197
FT STRAND 202 202
FT HELIX 213 216
FT STRAND 219 221
FT STRAND 223 224
FT STRAND 226 226
FT TURN 235 236
FT STRAND 242 242
FT STRAND 245 245
FT HELIX 246 248
FT TURN 249 252
FT HELIX 255 260
FT STRAND 263 267
FT TURN 268 270
FT STRAND 273 278
FT STRAND 287 287
FT TURN 289 291
FT STRAND 292 292
FT STRAND 295 307
FT STRAND 314 331
FT STRAND 337 337
FT STRAND 348 351
FT TURN 355 355
FT STRAND 363 363
FT TURN 366 367
FT STRAND 379 380
FT STRAND 382 382
FT HELIX 383 387
FT TURN 388 388
FT STRAND 391 392
FT STRAND 397 397
FT TURN 399 403
FT HELIX 405 408
FT STRAND 410 413
FT STRAND 423 426
FT TURN 429 431
FT TURN 433 437
FT HELIX 439 444
FT TURN 445 446
FT STRAND 447 451
FT STRAND 454 460
FT TURN 464 465
FT STRAND 467 467
FT STRAND 475 475
FT HELIX 485 489
FT TURN 490 490
FT STRAND 492 497
FT STRAND 503 508

```

FT STRAND 517 518
FT HELIX 523 525
FT STRAND 529 534
FT STRAND 539 539
FT TURN 542 543
FT STRAND 547 556
FT TURN 558 559
FT STRAND 561 565
FT STRAND 600 600
FT TURN 601 602
FT STRAND 604 604
FT STRAND 608 608
FT TURN 620 621
FT STRAND 622 623
FT HELIX 625 627
FT HELIX 635 637
FT TURN 638 638

```

Query Match 2.5%; Score 101; DB 1; Length 2208;

Best Local Similarity 19.8%; Pred. No. 32; Mismatches 214; Indels 176; Gaps 29;

Matches 114; Conservative 72; Mismatches 214; Indels 176; Gaps 29;

```

QY 10 NDGKLVYH-----GKTLTGVPDNNVLTPEG--SGRGV 40
DB 1662 NDGVLIVNTSKYPNNYVPGAVTEGQYLNLGGROTARTLMYNPTRBAGCGGVTCTGKV 1721
QY 41 TGAFAVATASHSKSLHVFPMGLBGLRMCCRFKLMNTQRMGTGGRDVFLEQPMLE 100
DB 1722 IGHNVGNGSH-----GFAAL-----KRSTFTSQG-----ETQWR-P 1755
QY 101 SKSESTDGENSPILYTVLLPYLE-GQFRAVLQGNDKNEIEICSGDNAVEFDGLHMY 159
DB 1756 SKS-----VGPRLNAPSKTKLEPSAFHYEGVKEPAV---LKNDRPLKTDDEEALFS 1807
QY 160 MHAGTPEFVINQAKAVAKEMQTLNREKKRLPSCLEDFEGWCTDAFY-TDVTAGVEE 218
DB 1808 KYVGNKITEVDKMEADYHAGLSLDINTEQCL-----DAUYGTD----- 1852
QY 219 GLKSL-----SQGSTPPEFLIIDGMOIENK-----AKDATECLVQGAQFATRLGKENT 271
DB 1853 GLEADLSTAGYR--YVAMGKKRDLINKQTRDKENQKLDITDYGILPL-VTVVDEL 1909
QY 272 KPOKTIQNEQMSGLKHLVHGAKOHNNKNNYVWHALAGYGVKPAATGMEHTDALAY 331
DB 1910 RSKTYVE-----QGSRLLEASLNDV-----AMTAFGNTYAAF--H----- 1946
QY 332 PVQSPGVL-----GNQPTVMDSLAVHGLGVHPKKVNF-YNELHAYLASCVDGYVD 385
DB 1947 --KNRQVITGSAVGCDDPDLFWFSKIPY-----LMEKILFAFDYTGVDASLSAFMEALKMV 1999
QY 386 VQNIETLGAGHGRVSLTRSYHNALEASINFTDNCIACMCHNTDGLYSARQTAIVR 445
DB 2000 LEKI-----GFDVRYDYIDYLNHSHNLKKNKYCVKGMSPSCSGSIFSMINNLIR 2053
QY 446 A-----SDPF--YPRD-----PASHIHLSV 465
DB 2054 TLLKTYKCIDDLKMLAYGDDVYASTPHEVDASLSLAQSGKDYGLTWTPRADKSANFEVY 2113
QY 466 AY-NSLIFGEFQMPDWD--MFHSLRPADYHAAR 497
DB 2114 TWENVTEFKRFRRADEKYPFLIHVPMKKEIHESIR 2149

```

RESULT 14

DYHC_ANTCR STANDARD; PRT; 4466 AA.

AC P39057;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

OS Dynein beta chain, ciliary

OS Anthracidaris crassispina (Sea urchin)

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

```

OC Echinoidae; Euechinoidae; Echinacea; Echinoidae; Echinometridae;
OC Anthracidaris.
OX NCBI_TaxID=7629;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91326104; PubMed=1830928;
RA Ogawa K.;
RT "Four ATP-binding sites in the midregion of the beta heavy chain of
  dynein."
RL Nature 352:643-645(1991).
CC -!- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
  FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
  DYNEIN HAS ATPASE ACTIVITY.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
  THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: D01021; BA00827.1; -.
DR PIR: S17231; S17231.
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: PF03028; Dynein_heavy; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
  Coiled coil.
FT 482 502 COILED COIL (POTENTIAL).
FT 627 643 COILED COIL (POTENTIAL).
FT 734 805 COILED COIL (POTENTIAL).
FT 1036 1056 COILED COIL (POTENTIAL).
FT 1306 1337 COILED COIL (POTENTIAL).
FT 1443 1468 COILED COIL (POTENTIAL).
FT 1950 1978 COILED COIL (POTENTIAL).
FT 3033 3134 MICROTUBULE-BINDING (POTENTIAL).
FT 3263 3325 COILED COIL (POTENTIAL).
FT 3573 3642 COILED COIL (POTENTIAL).
FT 154 161 ATP (POTENTIAL).
FT 1852 1859 ATP (POTENTIAL).
FT 2133 2140 ATP (POTENTIAL).
FT 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
SO SEQUENCE 4466 AA; 511772 MW; C465CC5C6C60325D CRC64;

```

Query Match 2.5%; Score 101; DB 1; Length 4466;

Best Local Similarity 19.5%; Pred. No. 89; Mismatches 289; Indels 222; Gaps 39;

Matches 147; Conservative 95; Mismatches 289; Indels 222; Gaps 39;

```

QY 37 RGLVYGAFAFGATASHSKS-----LHVPFMGLBGLRMCCRFKLMNTQRMGTGGRDVP 91
DB 1670 RSTVRSQFADAVSYEEKRPREQWLYDPAQY-----ALATTQVWMTTE----- 1712
QY 92 LETQPMLES--KSESTDGENSPIT-YTVLLPYLEGQFRAVLQGNDKNEIEICSGDNA 148
DB 1713 VNISFARLEEGHNSMKDYNNKKQILQNLTLGLLGK--LTKDRQKIMTIC----- 1762
QY 149 VETDGLHMYVWNAQTNFEVLNQAUKAVEKNMOTFLNREKKRLPSCLDLDFG--WCT 203
DB 1763 -----TIDVHAR-----DVVAMMVLKKVDSQAF-----QHSQLRHRMAD 1798
QY 204 WDAF-YTVDTAGVEEGLKLSLQSTPPEFLI--IDGMOIENKAKADATEC----- 252
DB 1799 DDKHGYANICDAQFQYSEYEL--GNPRLVITPELTD-----RCYITTLQSL 1842
QY 253 -LYQSGAQAFAIRLTGKIKNTKFGOKKLQNEQMSGLKHLVHGAKOHNNK--NYVWHAL 308
DB 1843 HLVNMGARAGPAGTGKTEYTK-----DLGRALGIMVYVNGSGEOMDKSCGNITYGSLAO 1896

```



```

OY 309 AGWGWKPAATGMEHYDATALAYVQSPGLGNPDIVMSLAVHGLVHPKRVENFN 368
DB 1897 TGAMG-----CFDEFNLSVEVLSSVAVQKCVQDAR-----DKKERNFNG 1939
OY 369 ELHAYLASGVGVKVDVONITETGAGHGVSILTRSYHALLASIASNFTDNCIACM 428
DB 1940 EELSLIPSVG-----IFITMPPGYAGRTLEPENL-KALFRPCAMVVPDFELICEI 1988
OY 429 CHNTDGLYSAKOTA-----IVRASDPFYRDPASHIITHISSAYNSLFGEMQ 477
DB 1989 MLYABEFLKARLLARKFTITLYLCKELLISKODHY--DWGLRAIKSVLVVAGSLKRGDQR 2046
OY 478 PWDMEHSHIPADYHAAARAIGCPIYVS-----DKPGNHFDLTK----- 519
DB 2047 PEDQVLT--NRALRDFNVPRKIVSDTPFVEMGLIGDLEPALDVRRRRDDEKVKVOSTIDL 2104
OY 520 KLVLPDGSYLK-AQLPG-RPTDLSLFV--DPAKRTSLIKTINLKKSGVGVENCQAG 575
DB 2105 KLAQEDSEFLKVVQLEELLAVRHVSFVIGMAGTGRKSOVLKV--LNKT----- 2149
OY 576 WCKIEKTRIDHDSPTGLT-----ASVCASDVLLITQVAGAEWL-- 614
DB 2150 YSNMRKRPVFDLNRKATYNDLFGIINPATREKMDGLFSVLMRMSNITH--DGKMWLVL 2208
OY 615 -GD-----TIVAYRSGEVIRLPKGVSIPTLVK-LEFELHPCPIQELAP-SISFPA 664
DB 2209 DDDIDPMWIESINTVMDNKNKVLTLASNERIPTLPSMRILEISH--LKTATPAVVSAG 2265
OY 665 IGLDMFNVTG-----GAVEQVEIHNRATKTI 691
DB 2266 ILYINPSDLGMPIYTSWIDREVOSERANLTI 2298

RESULT 15
ID DYNC TRIGR STANDARD: PRT: 4466 AA.
AC P23098:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dynein beta chain, ciliary.
OS Tripleneustes gratilla (Hawalean sea urchin).
CC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidae; Euechinoidae; Echinacea; Temnopleuroidea; Tokopneustidae; Tripleneustes.
OX NCBI_TaxID:7673;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Blastula;
RX MEDLINE-91326103; PubMed-1830927;
RA Gibbons I.R., Gibbons B.H., Mocz G., Asai D.J.;
RT "Multiple nucleotide-binding sites in the sequence of dynein beta heavy chain.";
RL Nature 352:640-643(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92020893; PubMed-1833761;
RA Gibbons I.R., Asai D.J., Ching N.S., Dolecki G.J., Mocz G., Phillips C.A., Ren H., Tang W.Y., Gibbons B.H.;
RT "A PCR procedure to determine the sequence of large polypeptides by rapid walking through a cDNA library.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8563-8567(1991).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES. DINEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA), THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59603; CAA42170.1; -.
DR PIR: S17653; S17653.
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: Pf03028; Dynein_heavy; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
KW Coiled coil.
FT DOMAIN 627 643 COILED COIL (POTENTIAL).
FT DOMAIN 733 805 COILED COIL (POTENTIAL).
FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
FT DOMAIN 1950 1978 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3033 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
FT NP_BIND 154 161 ATP (POTENTIAL).
FT NP_BIND 1852 1859 ATP (POTENTIAL).
FT NP_BIND 2133 2140 ATP (POTENTIAL).
FT NP_BIND 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
FT VARIANT 611 615 MISSING.
FT VARIANT 3356 3358 LPS -> LITGNFCCFMTAG.
SQ SEQUENCE 4466 AA; 511771 MW; 2A695BF836911E CRC64;

Query Match 2.5%; Score 101; DB 1; Length 4466;
Best Local Similarity 19.3%; Pred. No. 89;
Matches 150; Conservative 94; Mismatches 264; Indels 286; Gaps 42;

OY 37 RGLVTGAFVATYSHSKS-----LHVPKGVLEGLRMCGRFKLMMTQRMGTGGRVP 91
DB 1670 RSTYRSQFADAVVSYEKPREQWLDYPAQV-----ALATQVWMTTEVINSFAR-- 1719
OY 92 LETQFMLESKESETPDQENSPIT--TYVLLPILLEGOFRAVLQGNKNEIEICLESDDNAV 149
DB 1720 -----LEBEHESMADYKQKIQQLNTLGLLIG--LTKDROKIMTIC----- 1762
OY 150 ETDGLHMYMAGTNPFEVINOAVKAVEKNQTFLEKREKRLPSCLDMFG-----WCTW 204
DB 1763 -----TIDVHAR--DVAVMMVLKVDNAQAF-----QVLSQRLHNMAD 1799
OY 205 DAF-YTIDVTAEGVEBGLKSLSSQGTTPRPLI--IDGWOQIENKAKADATPC----- 252
DB 1800 DKHCYANICDAQFKSYEYL--GNTPRLVITPLTD-----RCYITLTQSLH 1843
OY 253 LVQGAOFATRLTGIRENTEFKOKLONNEDMSGKHLVHGAQONHNVK--NVYVWHALA 309
DB 1844 LVMSGAPAGAGKGTETTK-----DLGRALGIMYVNCSEQMDYKSCGNITYGLSOT 1897
OY 310 GYGVGVKPAATGMEHYDATALAYVQSPGLGNPDIVMSLAVHGLVHPKRVENFNE 369
DB 1898 GAGG-----CFDEFNLSVEVLSSVAVQKCVQDAR-----DKKERNFNGE 1940
OY 370 LHHAYLASGVGVKVDVONITETGAGHGGRVSLTRSYHALLA----- 413
DB 1941 EISLIPSVG-----IFITMPPGYAGRTLEPENL-KALFRPCAMVVPDFELIC 1986
OY 414 -----SIASNFTDNGCIA-----CMC-----HNTDGLYSAKOTATYVRAS--DDFY 451
DB 1987 EIMLVABEFLKARLLARKFTITLYLCKELLISKODHYDWGLRAIKSVLVVAGSLKRGDQR 2046
OY 452 PRDP-----ASHIITHISSAYNS--LFLGFMOPDMDMFHSHIPADYHAAARAIGCPIY 505
DB 2047 PEDQVLMRALRDFNVPRKIVSDTPFVEMG-----LIGDLEPALDV----- 2085
OY 506 VSDKPCGNHNFDDLK-----KVLVDPGGSYLK-AQLPG-RPTDLSLFV--DPAKRTS 552
DB 2086 -----PRRMDPEKVKVOSTIDLKLAQEDSEFLKVVQLEELLAVRHVSFVIGMAGTGRKSO 2141

```

```
OY 553 LKTNLNCGSGVGVFNCQAGWCIEKKRIHDSPTLT----- 594
Db 2142 VLKV--LNKT-----YSNMKRKPVLDLNPKAATNDELFGIINPATREKDG 2186
OY 595 -ASVCASDVDLITQVAGAEWL--GD-----TIVYVRSGEVIRLPKGVSIPTVLKV- 642
Db 2187 LFSYIMRDMNITH-DGPKWIVLDGDIDPMMIESLNTVMMDNKKVLTLASNERIPLTFPSMR 2245
OY 643 LEFELFHPCPIQETAP-SISFAIGLDMENTG-----CAVEQVEIHNRAATKI 691
Db 2246 LFFEISH--LKTATPATVSRAGILYINPSDLGWNPIVTSWIDTREVQSERANLTI 2298
```

Search completed: June 4, 2002, 09:10:09
Job time: 236 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 09:05:38 ; Search time 34.3 Seconds

(without alignments)
3782.688 Million cell updates/sec

Title: US-09-810-186-1
Sequence: 1 MVTVPKISVNDGKLVHNGKT.....TFSTIPVSPPEMYRMSIEIQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.19:**

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3219.5	80.6	773	10	Q9SCM1
2	3215.5	80.5	773	10	Q94A08
3	3181.5	79.6	773	10	Q43408
4	3158.5	79.0	773	10	Q9MAM7
5	2951.5	73.9	757	10	Q40077
6	1915.5	47.9	1170	10	Q40077
7	1732	43.3	1170	10	Q9M422
8	1422.5	35.6	783	10	Q9M422
9	1394	34.9	784	10	Q9M422
10	1348.5	33.7	816	10	Q94276
11	1307	32.7	807	10	Q94276
12	1284	32.1	857	10	Q94276
13	1228.5	30.7	357	10	Q93466
14	1228	30.7	357	10	Q93466
15	986.5	24.7	283	10	Q9FWM2
16	753.5	18.9	371	10	Q04607

17	517	12.9	702	2	Q93CM6	Q93cm6 bifidobacte
18	442.5	11.1	649	17	Q97094	Q97094 sulfolobus
19	437.5	10.9	674	17	Q96XG2	Q96XG2 sulfolobus
20	410	10.3	125	10	Q42099	Q42099 arabidopsis
21	166.5	4.2	88	10	Q42092	Q42092 arabidopsis
22	126	3.2	657	16	Q83140	Q83140 treponema p
23	120.5	3.0	2602	4	Q95369	Q95369 homo sapien
24	120.5	3.0	2602	4	Q95369	Q95369 homo sapien
25	119.5	3.0	1715	10	Q92V68	Q92V68 lactuca sat
26	116	2.9	1732	5	Q9VEZ2	Q9VEZ2 drosophila
27	114.5	2.9	1447	10	Q9C7H8	Q9C7H8 arabidopsis
28	114	2.9	4545	11	Q61291	Q61291 mus musculu
29	112.5	2.8	4545	11	Q920Y4	Q920Y4 mus musculu
30	112.5	2.8	4545	11	Q912X7	Q912X7 mus musculu
31	111	2.8	953	16	Q9A7G9	Q9A7G9 caulobacter
32	110	2.8	905	10	Q9A7K3	Q9A7K3 arabidopsis
33	109	2.7	684	2	Q9L1F8	Q9L1F8 streptomyce
34	109	2.7	1499	10	Q9LHM4	Q9LHM4 arabidopsis
35	108	2.7	2658	13	Q90WFO	Q90WFO gallus gall
36	106.5	2.7	500	10	Q9SHK2	Q9SHK2 arabidopsis
37	105.5	2.6	780	13	Q9PUI4	Q9PUI4 xenopus lae
38	105	2.6	775	2	Q9X6M6	Q9X6M6 enterococcu
39	105	2.6	1601	16	Q928Z7	Q928Z7 listeria in
40	104.5	2.6	3674	5	Q9TW65	Q9TW65 caenorhabdi
41	104.5	2.6	3674	5	Q9TYG9	Q9TYG9 caenorhabdi
42	104	2.6	754	10	Q9LEF1	Q9LEF1 arabidopsis
43	104	2.6	1669	10	Q9LXK4	Q9LXK4 arabidopsis
44	104	2.6	3895	12	Q9P297	Q9P297 pestivirus
45	103.5	2.6	2212	12	Q91DD4	Q91dd4 ebola virus

ALIGNMENTS

RESULT 1
Q9SCM1 ID Q9SCM1 PRELIMINARY; PRT; 773 AA.
AC Q9SCM1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE IMBIBITION PROTEIN HOMOLOG.
GN TBH10.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoorge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Queller F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133248; CAB66109.1; -;
SQ SEQUENCE 773 AA; 85143 MW; 0852F9E67952C8D3 CRC64;

Query Match 80.6%; Score 3219.5; DB 10; Length 773;
Best local similarity 76.6%; Pred. No. 3.6e-258;
Matches 593; Conservative 73; Mismatches 83; Indels 25; Gaps 3;

QY 1 MVTVPKISVNDGKLVHNGKTILNGVDPNVLLPGSGRGVGTAFVGAATASHKSLHVFPM 60
Db 1 MVTVPKISVNDGKLVHNGKTILNGVDPNVLLPGSGRGVGTAFVGAATASHKSLHVFPM 60
QY 61 GVLGRLRMCCEFRKLMNTORWGSGCKDIPLETOMLESKDEVEGNDADPFTYVFL 120
Db 61 GVLGRLRMCCEFRKLMNTORWGSGCKDIPLETOMLESKDEVEGNDADPFTYVFL 120

```

QY 120 PLEGGFRAVLQGNKNEIEICLESNDNAVETDOGLHMYVMHAGTNPFEVINOAVKAYER 179
D 121 PLEGGFRAVLQGNKNEIEICLESNDNAVETDOGLHMYVMHAGTNPFEVINOAVKAYER 180
QY 180 HMOFTLHREKKRLPSCGLDMFGMCTWDAFYTDYTAESVEEGKLSLSGGTTPREFLLIDGG 239
D 181 HMOFTLHREKKRLPSCGLDMFGMCTWDAFYTDYTAESVEEGKLSLSGGTTPREFLLIDGG 240
QY 240 QOENKAKDTECLVDEGAOFATRLTGKENTFKOKLONNOMSGIKHLVHGAOKHNV 299
D 241 QOENKAKDTECLVDEGAOFATRLTGKENTFKOKLONNOMSGIKHLVHGAOKHNV 299
QY 300 KNYVVMHLAGYGVGKPAASGMEHYDATALAYVPOSPGVGNOPDIYMDSLAVHGLGLVH 359
D 300 KNYVVMHLAGYGVGKPAASGMEHYDATALAYVPOSPGVGNOPDIYMDSLAVHGLGLVH 359
QY 360 PKRYVNFNEHSLAYASGIDGVKVDYONITETLAGIGRGVSLTRSYOQALASIANRF 419
D 360 PKRYVNFNEHSLAYASGIDGVKVDYONITETLAGIGRGVSLTRSYOQALASIANRF 419
QY 420 TDNGCIACMCHNTDGLYSAKOTAIYRASDDFYPRDPASHTIHISVAYNSLFLGEFMOPD 479
D 420 TDNGCIACMCHNTDGLYSAKOTAIYRASDDFYPRDPASHTIHISVAYNSLFLGEFMOPD 479
QY 480 WDMFSLHPADYHAARAAGCPIYVSDKPGNHNFDLLKLVLPDGSVLRQALPGRPTR 539
D 480 WDMFSLHPADYHAARAAGCPIYVSDKPGNHNFDLLKLVLPDGSVLRQALPGRPTR 539
QY 540 DCLFADPARDGLISLTKIMNNKFTGIVGVNCGAGCKETKKNQIHDTSPGTLGSIRA 599
D 540 DCLFADPARDGLISLTKIMNNKFTGIVGVNCGAGCKETKKNQIHDTSPGTLGSIRA 599
QY 600 SDVDLITOVAGAEMIGDTIYVAYRSGEVIRLPKGVSLPYTLKYLEFELHFCPIQELIAPS 659
D 600 SDVDLITOVAGAEMIGDTIYVAYRSGEVIRLPKGVSLPYTLKYLEFELHFCPIQELIAPS 659
QY 660 ISFAIIGLDMFNTGAVEOVEIH-----NRAATKTIALSVR 696
D 660 ISFAIIGLDMFNTGAVEOVEIH-----NRAATKTIALSVR 696
QY 697 GGRGREGYSSORPLKCVYGAETDFNDSFGLTTFSTIPVSPDEMYRMSIEIOV 750
D 697 GGRGREGYSSORPLKCVYGAETDFNDSFGLTTFSTIPVSPDEMYRMSIEIOV 750
QY 720 GCGRGYASSORPLKCAVESTETDFTYDAEVLVTNLNPTREEMFWMHVEILV 773
D 720 GCGRGYASSORPLKCAVESTETDFTYDAEVLVTNLNPTREEMFWMHVEILV 773

RESULT 2
ID Q94A08 PRELIMINARY; PRT; 773 AA.
AC Q94A08;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE PUTATIVE INHIBITION PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamuya A., Kallin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Saito R.W., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shiozaki K., Davis R.W., Ecker J.R., Theologis A.,
RA "Full length cDNA of gene T8H10.120/AT3957520 (GI:6706422)."
RA Submitted (Jul-2001) to the EMBL/Genbank/DBD databases.
DR EMBL; AT050772; AK92707.1;
RN NCBI_TaxID=3702;
RP SEQUENCE 773 AA; 85171 MW; 18BDF9E67952C801 CRC64;

```

```

Query Match 80.5%; Score 3215.5; DB 10; Length 773;
Best Local Similarity 76.5%; Pred. No. 7.7e-258;
Matches 592; Conservative 74; Mismatches 83; Indels 25; Gaps 3;

QY 1 MPTVKRISVNDGKLYVHGKTLTGVDNVVILPGSGRLVTGAFVGAFAASHKSLHVP 60
D 1 MPTVKRISVNDGKLYVHGKTLTGVDNVVILPGSGRLVTGAFVGAFAASHKSLHVP 60
QY 61 GYLEGIRFMCCEFRKLMWTOHRRGSGKIDPLETQFMLESDEVEGMDDDAPYTYTFL 120
D 61 GYLEGIRFMCCEFRKLMWTOHRRGSGKIDPLETQFMLESDEVEGMDDDAPYTYTFL 120
QY 120 PLEGGFRAVLQGNKNEIEICLESNDNAVETDOGLHMYVMHAGTNPFEVINOAVKAYER 179
D 121 PLEGGFRAVLQGNKNEIEICLESNDNAVETDOGLHMYVMHAGTNPFEVINOAVKAYER 180
QY 180 HMOFTLHREKKRLPSCGLDMFGMCTWDAFYTDYTAESVEEGKLSLSGGTTPREFLLIDGG 239
D 181 HMOFTLHREKKRLPSCGLDMFGMCTWDAFYTDYTAESVEEGKLSLSGGTTPREFLLIDGG 240
QY 240 QOENKAKDTECLVDEGAOFATRLTGKENTFKOKLONNOMSGIKHLVHGAOKHNV 299
D 241 QOENKAKDTECLVDEGAOFATRLTGKENTFKOKLONNOMSGIKHLVHGAOKHNV 299
QY 300 KNYVVMHLAGYGVGKPAASGMEHYDATALAYVPOSPGVGNOPDIYMDSLAVHGLGLVH 359
D 300 KNYVVMHLAGYGVGKPAASGMEHYDATALAYVPOSPGVGNOPDIYMDSLAVHGLGLVH 359
QY 360 PKRYVNFNEHSLAYASGIDGVKVDYONITETLAGIGRGVSLTRSYOQALASIANRF 419
D 360 PKRYVNFNEHSLAYASGIDGVKVDYONITETLAGIGRGVSLTRSYOQALASIANRF 419
QY 420 TDNGCIACMCHNTDGLYSAKOTAIYRASDDFYPRDPASHTIHISVAYNSLFLGEFMOPD 479
D 420 TDNGCIACMCHNTDGLYSAKOTAIYRASDDFYPRDPASHTIHISVAYNSLFLGEFMOPD 479
QY 480 WDMFSLHPADYHAARAAGCPIYVSDKPGNHNFDLLKLVLPDGSVLRQALPGRPTR 539
D 480 WDMFSLHPADYHAARAAGCPIYVSDKPGNHNFDLLKLVLPDGSVLRQALPGRPTR 539
QY 540 DCLFADPARDGLISLTKIMNNKFTGIVGVNCGAGCKETKKNQIHDTSPGTLGSIRA 599
D 540 DCLFADPARDGLISLTKIMNNKFTGIVGVNCGAGCKETKKNQIHDTSPGTLGSIRA 599
QY 600 SDVDLITOVAGAEMIGDTIYVAYRSGEVIRLPKGVSLPYTLKYLEFELHFCPIQELIAPS 659
D 600 SDVDLITOVAGAEMIGDTIYVAYRSGEVIRLPKGVSLPYTLKYLEFELHFCPIQELIAPS 659
QY 660 ISFAIIGLDMFNTGAVEOVEIH-----NRAATKTIALSVR 696
D 660 ISFAIIGLDMFNTGAVEOVEIH-----NRAATKTIALSVR 696
QY 697 GGRGREGYSSORPLKCVYGAETDFNDSFGLTTFSTIPVSPDEMYRMSIEIOV 750
D 697 GGRGREGYSSORPLKCVYGAETDFNDSFGLTTFSTIPVSPDEMYRMSIEIOV 750
QY 720 GCGRGYASSORPLKCAVESTETDFTYDAEVLVTNLNPTREEMFWMHVEILV 773
D 720 GCGRGYASSORPLKCAVESTETDFTYDAEVLVTNLNPTREEMFWMHVEILV 773

RESULT 3
ID Q43408 PRELIMINARY; PRT; 765 AA.
AC Q43408;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE PUTATIVE INHIBITION PROTEIN.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.

```


OY		1	MNYTPKISVNDCKIYVHGKTIILTIGVRDWNVLITPGSGRDLVTGATVGATASHKSCLHWPFM	60
Db		1	MTYAGACISTDSLDLVNGLHGVLEPENVLITPASGNALIDGAFIGVSDTGSHRFVSL	60
OY		61	GVLGEGLREMCRCRRFLMMWTOGMTCGCGADVPLETFQFMLESKE--SETDGEMSPITIVVL	119
Db		61	GKLIEDREFCVRFRLLMMWTOTGMGTNGKEICECTQFLLIVEANQSSDLJGGRDOSSSYVFL	122
OY		120	PLLEGCFRAVLGGNDKNEIEICLESSGDNAVEFDGLHMYHMAQTNPFEVINQAOKAV--	177
Db		121	PLLEGCFRAVLGGNANEIEICLESSGEDPTVDQFEGSHLIFVAASDPEDVITKAIVYMF	180
OY		178	-EKHNQTEFHREKRKKLPSCDLMFGMCTDAITYTDTYTAGVGEELK-----SL	223
Db		181	LKSQKLTSLCPNFEEFMRPMLNMFGMCTDAEYTVNTADVAKQGLESNODLTKPALILCSL	240
OY		224	SOGGTRPRRLIIDDDMOQIENKAKATECLIQEOGAOFATRLTGIKENTKFOKKLONNEOM	283
Db		241	KAGWVTPKFVIIIDDQWQSV-QMDETSVEFNADMANANRNLTHIKEHHKFKQDKEGEHRY	299
OY		284	S-----GLKHLYHGAKQHNNVKNVYVHAALAGYGVKAPRAATGMEHYDTALAIVPYOSPVL	339
Db		300	DDPSLSLGHVITDIKSNNSLSKYVYWAHALITGYMGVYKRGVSGMEHYESKAYPAVYSPGVN	359
OY		340	GNOPVIDSLIAVHGGLGVHKVKNFYNELHAFLASCQVDGVKVYDQNIETFLGAGHG	399
Db		360	SENEGCCLESTIRKMGSLGVNPEKVFSEFNDLHSLAVSGVDGVKVYDQNIETFLGAGHG	419
OY		400	RVSLTRSYVHAALLEASIASNFENDNCIACMCNHTDGLYSAKOTALIVRASDPFPRDPAHRT	459
Db		420	RVKLAKKYHQALEASISNPNPDNCIIISMINTDGLSAKKTATIRASDDPFWRDRASHT	479
OY		460	IHISSVANSLFLEFPQPMDFHSLSHPADYHAAARAIGGCPIIYVSDKPGMHNFDLK	519
Db		480	IHIASSVAANTLFLEEFQMPPMDMFHSLSHPAEMAHAARAVGCAIYVSDKPGQHDFFLLR	539
OY		520	KLVIPDGSVTLAOLIPGRTRTSLSLVDPARDRTSLKIMNLNKSCGVYGVNCGAGACHKI	579
Db		540	KLVLRDGSILIAKLIPGRTR-----	559
OY		560	EKKTRIHDTSPTLTASVCASDVULLTVQVAGAEWLGDTIYVARSGEVIRLPKGSVLPYT	639
Db		560	-----ELVYLPRDTSLPYT	573
OY		640	LKVLEFELFHCPQIOELAPSISFAPIGLDMENITGCAVEOVEIHNRAAKTTALSVRGG	699
Db		574	LMPREYEFTVYPAPKEFSDSGSKFAVGLMEMFNMSGAIYSLRYDDGCTKPVVMKLRGSG	633
OY		700	REGVYSS-ORPLELCVVGGAETDF 721	
Db		634	LVGYVSSVRPRRSVTVDSDDVEY 656	
RESULT		7		
Q9M442				
Q9M442	PRELIMINARY;	PRT:	386 AA.	
AC	Q9M442;			
DT	01-OCT-2000 (TREMBLrel_15, Created)			
DT	01-OCT-2000 (TREMBLrel_15, last sequence update)			
DT	01-JUN-2001 (TREMBLrel_17, last annotation update)			
DE	PUTATIVE IMBIBITION PROTEIN (FRAGMENT).			
OS	Cicer arletium (Chickpea) (GARBANZO).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;			
OC	eumossids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.			
OX	NCBITaxId=3827;			
RP	[1]			
RP	SEQUENCE FROM N.A.A.			
RC	STRAIN=CV. CASTELLANA; TISSUE=ETOLATED OSMOTIC STRESSED EPICOTYLS;			
RA	Dopico B., Romo S., Labrador E.;			
RT	"A putative imbibition protein is expressed in chickpea epicotyls.";			
LU	Submitted (JAN-2000) to the EMBL/Genbank/DDbj databases.			

```

DR EMBL: AJ271668; CAB71135.1; -.
FT NEM_TER 1
SQ SEQUENCE 386 AA; 42180 MW; 357AF9FB8D71650 CRC64;

Query Match
Best Local Similarity 43.3%; Score 1732; DB 10; Length 386;
Matches 327; Conservative 13; Mismatches 22; Indels 24; Gaps 1.

QY 389 IETTGAGHGGRVSLTRSYHHALEASIASNFTDNGCIACMCHNTDGLYSAKOTAIYRASD 448
DB 1 IETTGAGHGGRVSLTRSYHHALEASIANRPNADNGCIACMCHNTDGLYSAKOTAIYRASD 60
QY 449 DEPRDPAHTTHISSVAANSFLFGFMDPDMDFSLHPADYHAANAIGGCPITYSD 508
DB 61 DEPRDPAHTTHISSVAANSFLFGFMDPDMDFSLHPADYHAANAIGGCPITYSD 120
QY 509 KRGNNHFDLLKLVLPDGSVLRALQRLGPRTRPSLFVDPARDRTSLIKTNLKSGGVGV 568
DB 121 KRGNNHFDLLKLVLPDGSVLRALQRLGPRTRPSLFVDPARDRTSLIKTNLKSGGVGV 180
QY 569 FNCQAGWCKIEKTRIRHDTSPGTLTASVCSADVLTITVAGAEWLGDTIYVARSGEYI 628
DB 181 FNCQAGWCKVEKTRIRHDTSPGTLTSSVSASDVQINQVAGVEWHGERTIYVARSGEYI 240
QY 629 RLPGKGSIVPTKLVEFELFHCPQIOLAPISFPAITGLDMFNPGAAVEYEH----- 683
DB 241 RLPGKGSIVPTKLVEFELFHCPQIOLAPISFPAITGLDMFNPGAAVEYEHKASDN 300
QY 684 -----NRAATKTIASLVGRGPRFGYSSORPLKCVAGAEITDPPND 724
DB 301 KQELFDGEVYSELTLISLSPNRTKTNATVALKAVGSKFGYSSQHPQLQCAVDGIDTDPND 360
QY 725 SETGLTFSIPVSPEDMTWMSIEIOV 750
DB 361 SETGLTFSIPVPOEGMYRWSIEIOI 386

RESULT 8
Q9FND9 PRELIMINARY; PRT; 783 AA.
AC Q9FND9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE RAPINOSE, SYNTHASE, PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL: AB006702; BAB11955.1; -.
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D787188BAC CRC64;

Query Match
Best Local Similarity 35.6%; Score 1422.5; DB 10; Length 783;
Matches 297; Conservative 142; Mismatches 280; Indels 49; Gaps 17;

QY 4 TPKISVNDGKLVVGGKTLNGVDPNNVLT-----PGSGRGLVTAFAVGATA-SISK 54
DB 19 TEKRLIEDSTILANGQVVLIDVYVNVYLTSPYLVKDGVPLDVASQSFGEFNIDGEPK 78

```

QY	55	LHNPRGVLLEGRMCCFRKIKMMWMTQRMOTGCRDRLPTLEPFLTESKSEI--DENSP1	113
Db	79	HNVAISIKLRINRMSIFERKVMWTFTHWGSNGRDLENELQIILLDSSGSDSGPGSGSR	138
QY	114	IYVILLPLEGQFAVLQAGDKNEIEICLSSGNAVETDGLHMYMAGNTNPFVEYINOA	173
Db	139	PYVILLPLBESFSSPSFGSEDDDAVACVBSGTEVTGSEFRRQYIVYHAGDDPFLKYDA	198
QY	174	VKAVEKHQVFLNHEKKRLPSCLDMFQWCIMDAFYTDYTAEGVEDEKLKLSLQSGPPRRPL	233
Db	199	MKVIRVHNMTFKLLEEKSPSPRIDYDKFQWCIMDAFYTLVNDGVNKKVCLLVDSGCGPGLV	258
QY	234	IIDDGWOOLEKAKDAREC-----LVDSGAQFATRLTGIKENTGYFOKKLQNNQMS--GLKH	288
Db	259	LIDDMQSIGSDS--DGLDVEGEMNTVAGDEQMPCKLLKFEENHKFKYVSPKQDNDVGMA	317
QY	289	LVHGAK-QHNHKNVYVMHALAGVSGVCPKPAQTOMEHYDTALAPVQSPGVLGNODIYM	347
Db	318	FVRDLKDEFSYVDYIYVHMLACGVCGLRPEADALP--PSTIIRPELSPBLKLTMDLAV	375
QY	348	DSLAVHGLVLPKPKVFNFVNEELHAYLASGVDSGVQVDVQNIIEETGAGHGVRSLRSY	407
Db	376	DKIIEGTIGFASPLAKEEYEGLSHLQNGSIDCVKDYHILEMLCQKTKGGRVLDLAKY	435
QY	408	HHLEASIASNFTNGCICACMCNTDGLY-SAKOTATVRASDDFYPRPAS-----	457
Db	436	FKALTSVNMKHFNGGVYASMEHNDMPFLQTEAISLGRVDGDPWCTDPGSDNGTFWLQ	495
QY	458	-HTTHISSVAINSLFLGEFQNPQMDHFSHPADYHAAARAAGCGPIYVSPKQNNH	515
Db	496	GCHMHW---CAYNSLMMGNF10PQWDMFQSTHPCAEFHASRAISGGPIYISQVGNHDE	552
QY	516	DLKRLKLYPDSVLYLAQLPGRPTDLSVDPDARPORTSLKIMNLNKSQGVGVFNCQAG	575
Db	553	DLKRLKLYPDSVLYLAQLPGRPTDLSVDPDARPORTSLKIMNLNKSQGVGVFNCQAG	612
QY	576	WCKIEKTRIHDTSPGTLTASVCAVDLIT-----QVAGAEMLGDTIYVAKSGEYIRL	630
Db	613	WCRETRRMQCESECVNLTLTATTSPPDYMMNMGSSPSISIANVE--EPALPLSOSKILL	669
QY	631	PGVSLPIYTLKLELHPHCPIQEI-APSISFAIGLDMFNTGAGVEQVEIHNRATK	689
Db	670	GLNDELLELPEFKELTTSVPVYTIENSNRPAPIGVNMNLNSGAIRSL-VYN---DE	725
QY	690	TALSVRGKRGKRGVYSQRPKCYVVGAEATDFNYDSETGLTTSIPVS	737
Db	726	SVGVGFGAGEPRYVASKKPVCSCLIDGVEVEFGY--EDSNVMQVQWPS	771
RESULT	9		
ID	092T62	PRELIMINARY:	PRT: 784 AA.
AC	092T62;		
DT	01-MAY-1999 (Tremblrel. 10; Created)		
DT	01-MAY-1999 (Tremblrel. 10; last sequence update)		
DE	01-DEC-2001 (Tremblrel. 19; last annotation update)		
DT	RAFIINOSE SYNTHASE (EC 2.4.1.82).		
GN	RF5.		
OS	Cucumis sativus (Cucumber).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid 1; Cucurbitales; Cucurbitaceae; Cucumls.		
OX	NCBI_TaxID=3659;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LEAF;		
RA	Ohsumi C., Nozaki J., Kida T.,		
RL	Submitted (Juz-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF073744; ADD02832.1;		
DR	Transferrase; Glycosyltransferase.		
SW	SEQUENCE 784 AA; 86920 MW; 3806491P0908933 CRC64;		

Query Match	34.9%	Score 1394	DB 10	Length 784
Best Local Similarity	37.6%	Pred. No. 1,1e-106		
Matches 289	Conservative 143	Mismatches 285	Indels 52	Gaps 15
QY	8	SVNDGKLIVHAKGKILITLGPVDNVLTIPGSGRGL----	VTGAFAVGATASHKSLHVFPMGV	62
Db	26	AIDGSDPTVNGHSHFLSDVPENIVASPSPTSIDKSPVSGCFGVFADSEDPDSHHVNSIGK		85
QY	63	LEGIREMCCFRKLMWMTORMGTCGRDPLLEQFPLILSKSESEFDGNSPIITVTLPL		122
Db	86	LKIDIREMSIRFRVWMTTHMVRNGDLESEQIVILEKSSD----	GRP--YVFLPIV	138
QY	123	EGCFRATVJGNDKNEILEITLSEGDNAVEDQGLHVVYHNAGTNPFVYINQAVAKEMHQ		182
Db	139	EGFPRTSIQGDDEVDVCEVSSSKSVVDASRSMLYTHAADDPFALYKEMKIVRHNIG		198
QY	183	TFILREKRRLSPCLDWFMCWTMDAFYTDVYTAGVEEGKLSLGGSTPPRELLIDGMOQI		242
Db	199	TFRLLEKTEPRGLVDRFGMCTMDAFYLVHNGVIEGVRIYHLDGCGPGLVLLDDGWOIS		258
QY	243	ENKAKDATECLVOE--GAQFATRLTGIKENTKFO-----	KLQNDMGSKLILVIGAK	254
Db	259	GHDSDPTTEKGMNQTAGEGMPCRLIKFPEENKFRDYVNPRAKTVGRRAGQKMKAFIDELK		318
QY	295	-QHNNVKNVYVWIALAGYNGCVKPRATGMEHNTDLAVPYOSGVLGNQPIVMDSLAVH		353
Db	319	GEFTVYVWYVWIALAGYNGWGLRPQVPLP--EARIQVIVSPGLQMTMEDLAVDKIVLH		376
QY	354	GLGVHVRKRYVNFENELHAYLASGVGVKVDVONITETLAGHGRVSLTRSYHNALEA		413
Db	377	KVGLVPRPEKAEWTEGSLHMLEKVGDYKIVYHLEMLCSDYGGRVDLKAYYKAMTK		436
QY	414	SIASNTDNGCIACMCNHTDGLY-SAKQIYIVRASDDYPRDPAS-----	HTIH	461
Db	437	SINKHFKNQVYIASMEHCNDFMLTGEATLSLGVGDGDFCDDPSGDPNGTFWLOGCHAVH		496
QY	462	ISSVAVNSLFGEMODMDMFHSILPRADYHAAARAGICGPITYSDRGNNHNDLKL		521
Db	497	--CANDSLWMGNKIHFDWDMFQSTHPCAFHRAASHAISGGPIIYSDSVGKHNDDLKL		553
QY	522	VLPDGSYVRAOLPFRPRDSLFWDPARDRTSLKINLKKCSGVVGVNCOGAWCKIEK		581
Db	554	VLPDGSILRSTYALPLPRDCLFEDPLHNGMTMLINLKLKFLVIGAFNGOGGWCRETR		613
QY	582	KTRIHDTSPGLTHAVCSAVCDVLLT--QVAGAEMLODTIVYRSGEYIRLPKGVSLPYT		639
Db	614	RNOCFSQYSKRVTSKTPNKDIEHMSGENPISIGVTFYALYLOAKKLLSKPSQDDIDIA		673
QY	640	LKVLFEFLFHCPCQIE-IAPSISMAAGLLDMNTGCAVEOEHNNRAATKIALSVGR		698
Db	674	IDPFEFELITVSPYTKLIQTSLSHAPAGIVNMNTSGAIGQSDVYDDUS--SVEIGKVC		731
QY	699	GFGYVSSORPLKCVGGAETFDVDESETGLTFSPVPSPEEMYRMSLE		747
Db	732	GENRVFASKRPACRIDGEDVGEFYTDQ--MYYVQV-----	WPID	771
RESULT	10			
ID	Q942T6	PRELIMINARY;	PRT;	816 AA.
DT	01-DEC-2001	(TREMBLrel. 19, Created)		
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
GN	P0583G08.2.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OX	NCBI_TaxID=4530;			
RP	SEQUENCE FROM N.A.			

RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(Ga3) genomic DNA, chromosome 1, PAC
clone:PO583G08."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003282; BAB64768.1;
SQ SEQUENCE 816 AA: 89588 MW: B316EDF3566C5178 CRC64;

Query Match 33.7%; Score 1348.5; DB 10; Length 816;
Best Local Similarity 37.5%; Pred. No. 7.1e-103;
Matches 301; Conservative 116; Mismatches 288; Indels 97; Gaps 18;

QY 5 PRISVNDKLVYHGKTIILTVGPDNVLLTPGSG-----RGLVTGAFVGAATASHKSLHY 57
DB 26 PRFTLKGRDLAVDGHPELIDVAPANRLRPASTLVNSDVPAALAASTFLGFDPAKADNV 85
QY 58 FMAGVLEGLREKCCFRFKIMMTQMGTCGRDVPLETQPMLE---SKSETDGENSPIL 114
DB 86 VIGKLRDTRFMSIFRFKVMWTHVWGNGRDVENETQMILDOGSKSPT---GPRP 141
QY 115 YTVLLPLEGOFRAVLIQ-GNDKNEIEICSGDNAVEVDQGLHMYMAGTNPFEVINO 173
DB 142 YVLLRPYBGPFRACLESKADYHNVLESSTVRSVSAAYLHAGDPEDLVKDA 201
QY 174 VVAVEKHMOTFLHREKKRLPSCIDMFGMCTDAFTDYTAEGVEGLKSLSGGTPPRL 233
DB 202 MRYVRAHILGTFRLMEKTPRPIVDKFGMCTMDAFYLVKHPGSGVWGVRLADGCGPRPLV 261
QY 234 IIDDGMOOI---ENKAKATEL--VQSGAOFATRLTIKENTKFOKKLQNNQNSGLKH 288
DB 262 ILDDWOSTCHDDDLGSGAEGEMNTSAGEQMPCLIFOEYKFR-----EYKGMWG 315
QY 289 LVHGAK-OHHNVKNYVYMHALAGYGVKYPATGMEHDTALAVPYOSPGVLGNDPDTVM 347
DB 316 FYREKKAFPRVEGYVYVNHALOGTYWGLRPGARGLP--PAKVAPRLSPGLQRTMEDLAV 373
QY 348 DSLAVHGLVHPKVFNFYVYELNAYLASGVGVGVYDV----- 386
DB 374 DKIVNNGVLDPRRARRELVELGSHLQASGIDGVKVDVIRHKWILYKIKDNATERFVL 433
QY 387 -----QNTIELGAGHGGRVSLTSSYHHALEASTASNTDNGCIACCHNTD- 433
DB 434 KSEIYRLDCVLLDLENVCEYEGRELAAYRAGLIESYRHHNGSVTASMEHCNDF 493
QY 434 GLYSAKQTAIVRASDDFYRDPAS-----HTHISVAVNSLFLGEFNOQPDMD 482
DB 494 MLGTEAVALGRVGDDEFCSTDPGSDPGSTFWLQCGHNV---CAYNSLMMGAFTHPWDM 550
QY 483 FHSJLRADYHAARAIIGGCPITYSDKFGNINFDLLKKLVLPDGSVLAQLPGRPTDSL 542
DB 551 FOSTHPCAFHAAASRAVSGGVYSDAVGCHDFLLRLRLAPDGTIILRCERYALPTFRLCL 610
QY 543 FVDPARDTSLIKLWNLKRCGGVGVFNCOGAGWCKTEKTRHIDTSPGTLTAAVCASDV 602
DB 611 FADLHOKTKMLKIMNVKESGVIGARFCOGGMSREARRMCAAGSVPTAASPADV 670
QY 603 DLITQVAGAEWL-----GDTI-VYARSGEVIRLPKGVSIPTVLKLVLEFELHFHCPIOEI 656
DB 671 -----EMSHGGGGGDRFAVYFVEARKQLLRDESEVELLEPTYELLYVAPRAI 721
QY 657 AP---SISFAIGLIDMENNGAIVEYEHNRATKITIALSVGRGFRGYSSORPLKCV 713
DB 722 VSPELGIGFAITGLANMLNGGAOYGFDAARKDGDVAEVAVAGBMAVSSARPLCK 781
QY 714 VGAETDENYDSETGLTTFPSIP 735
DB 782 VNGDAERKY--EDGIYTVDP 801

RESULT 11
Q9SYU4 PRELIMINARY; PRT; 807 AA.

AC Q9SYU4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PUTATIVE RAFINOSE SYNTHASE OR SEED IMBIBITION PROTEIN.
GN T7B11.23 OR ATAG01970.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
OC eucosids II: Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Huang E.N., Nascimento L., de la Bastide M., Habermann K., Vil M.D.,
RA Preston R.R., Spiegel L.A., See L.H., Shah R., O'Shaughnessy A., Rodriguez M.,
RA O'Shaughnessy A., Rodriguez M., Shekher M., Swaby I., Schutz K.,
RA Parnell L.D., Dedhia N.N., McCombie W.R.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

QY 121 SEQUENCE FROM N.A.
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007138; AAD2659.1;
DR EMBL: AL161493; CAB80690.1;
SQ SEQUENCE 807 AA: 90122 MW: 963DCD5A827B338B CRC64;

Query Match 32.7%; Score 1307; DB 10; Length 807;
Best Local Similarity 37.3%; Pred. No. 1.9e-99;
Matches 290; Conservative 131; Mismatches 287; Indels 70; Gaps 18;

QY 8 SVNDKLVYHGKTIILTVGPDNVLLTPGSGRLVT-----GAFVGATAS 50
DB 35 NLESGSLCANDSPRILEDVQNTFPPSSHSTISDAPRLILRYQANHKGGFLGFTPE 94
QY 51 HSKSLHFPWGLVEGLREKCCFRFKIMMTQMGTCGRDVPLETQMLIESKSETDGEN 110
DB 95 SPSDRLTNSLGRFEDREFLSLFRFKMWSWAWIGKSGSDLAQETQVMMLKPEIDS--- 150
QY 111 SPITVYVLLPLEGOFRAVLOGNDKNEIEICSGDNAVEVDQGLHMYMAGTNPFEVY 170
DB 151 -----VVAIIPTEIGARFASLTPEKKNVILCAESGSTKVESEFKIAIHTICDNRYNM 206
QY 171 NOAVKAVEKHMOTFLHREKKRLPSCIDMFGMCTDAFTDYTAEGVEGLKSLSGGTPP 230
DB 207 KEAFSALRYHMTNFKLLEKKLPKIVDKFGCTWDACYLTVPAITWGVKFEFDGVCP 266
QY 231 RFLIIDDGMOOI---ENKAKATELVOGAOFATRLTIGIENTKFE---QKKLQNNQEM 283
DB 267 KEYIIDDGMOISTFDELDKDA-ENLVYLGGEOMTARLISFECKKFRNFKYKESLSGDV 325
QY 284 SG-----LKHVLGAKQHHNVKNYVYMHALAGYGVKYPATGMEHDTALAVPVQSPG 337
DB 326 SSGSMAAFTKDL---RLRFSISDDIYVMAHLCGAWGVAPER--MMDLAKAVAPELSPS 380
QY 338 VLGNDPDIYMDSLAVHGLVHPKVFNFYVYELNAYLASGVGVYDVYQNTIELGLAGH 397
DB 381 LGATMDLAVDKVVEAGIDLVHPSKAHEFYDSMHSLASVGTGAKIDVFQTLSEIAEHH 440
QY 398 GGRVSLTSSYHHALEASTASNTDNGCIACMCHNTDGLYSA-KQTAIVRASDDFYRDP- 455
DB 441 GGRVELAKAYYDGLTESMKTNKNGTDVYASMOQNEFFFLAKQIISIGRVGDDFWMDPY 500
QY 456 -----ASHTHISSVAVNSLFLGEFMOPDMDMFHSLRPADYHAARAIIGGCPY 505

```
Db 501 GDPGVYVWLOGVHICHS---YNSIMWGQMIQPDWDFQSDHVCALHNASRAITCGPVT 557
QY 506 VSDKPG--NNHFDLKLKLVLPDGSVLRALQPGHPTSLFVDPARDRTSLIKTMINLKCS 563
Db 558 LSDHLGASNNHFDLKLKLPDGDGIIPRCVHAYLPTRLSLKLPDSESLIKLFNNKKG 617
QY 564 GVGAVENCQAGWCKIEKTRIHDTSPGTLTASYCASDV--DLITQVAGAE--WLGDTIV 619
Db 618 GVIETFCNQAGAMSPREHREFKGYKECYTIVSGTVHVSIDIMDQNPBAASQVYTGDIYV 677
QY 620 YANSGEVIRL-PKGSVLPITLKLFEFLHPCIOELAPS-ISFAIIGLIDMFNQGAV 677
Db 678 YKQOSEELFENNSKSEAMKTLIPLSADLSFVPTLTVSSGVAFALGLINFCYGTV 737
QY 678 EGVETHNRAATKTALSVGRGRFVYSSQRPKCVVGAETDPNDYSETGLTTFSP 735
Db 738 QDMKV---TGDNSTIRVDYKGGKGRPMAYSSAPVKCYLNDKEAEKWEETGKLSFVFP 792

RESULT 12
Q9SBZ0 PRELIMINARY; PRT; 857 AA.
ID 09SBZ0:
AC 09SBZ0:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GALACTINOL-RAFTINOSE GALACTOSYLTRANSFERASE (EC 2.4.1.67).
OS Phaseolus angularis (Adzuki bean) (Vigna angularis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
  eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OC NCBI_TaxID=3914;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20117502; PubMed=10652123;
RA Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;
RT "Stachyose synthesis in seeds of adzuki bean (Vigna angularis):
  molecular cloning and functional expression of stachyose synthase."
RL Plant J. 20:509-518(1999).
KW EMBL; Y19024; CAB64363.1; -.
DR Transferase; Glycosyltransferase.
SQ SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;

Query Match 32.1%; Score 1284; DB 10; Length 857;
Best Local Similarity 35.1%; Pred. No. 1,7e-97;
Matches 296; Conservative 130; Mismatches 269; Indels 148; Gaps 20;

QY 9 VNDGKLVVHGKTIILGVDDNVLT-----PGSGRGLVT-----GAFVGATASHS 52
Db 23 LSDGKLVYKGVVLLSHVENVTFSSFSSTCYPRDAPSSILQRTVAASHKGGFLGFS---- 78
QY 53 KSLHVF-----MGVLEGLRFMCFFRKLMMWTQRMGTGGRDVPLETOFMLESKESK 105
Db 79 ---HSPDRLNLSGSRGRNFIILFRKTIWSTQWNGSGSDLOMTQWLLIEVPETE 135
QY 106 TUGENSPITTYVLLPLBSQFPAVYLOGNDKNEIEICLESSGNAYETDGLIMVYMHAGTN 165
Db 136 S-----YVVIIPILIEKFSFSAIHPGSDHVKICAESGSTQVASSFGAIAVYHAEV 187
QY 166 PREVINQAVKAEKHMOTFLHREKRLPSCLDPMFGCTWDATFYDTVTAEGVEGLKSLSQ 225
Db 188 PYNIMREAVSAIRVHIDSRLLIEKTVIRIYDKFGWCITWDATFYLVNVGVGHGKJDSSE 247
QY 226 GGTTPREFLIIDGWOQIENKAKDATE--CLVQEGAGPATRLTGIKENTKFOKKIQ---- 278
Db 248 GVAAPREFVVIDDQWQSVNDDDEDPNDAKNIIVIGSEOMTARLHREEDKFRKYQKGLL 307
QY 279 -----ANEQMSGL-----KHL-----YHSAK 294
Db 308 GPNAPSFNDETIKELISKIEAEHLGKQAAAISAGSDLAETELMIVKREIIDLFGSK 367
```

```
QY 295 -----OHNNKANYVVMHALAGVWGVPAAATGMEHY 325
Db 368 GKESENSGGCCCAAECCGAMDFTTDLTERGGLDLVYVMHALGCGWGVAPGTT---HL 424
QY 326 DTALAYPVQSPGVLGNODIYVDSLAVHGLIVHKKVFNFNELHAYLASCVDYKAVD 385
Db 425 DSKITPCKLSPGLVGTMMDLAVDKIVEGSIQIVHQAANDLYDSHSHYLAOTGYGVYID 484
QY 386 VONITETLGAGGGRVSLRSYHHALEASIANFNFDNCIACMCNNTGGLY-SAKQTAIV 444
Db 485 VHSLEYVCEEYGVHETAKAVYDGLVNSTIKNENGSGIISMQOCNDPFELGTQIDFG 544
QY 445 RASDPEPRDP-----ASHTIHSSVAVNSLFLGERMOPDWMDFSLHPADYH 493
Db 545 RYGDPEWEDPPMGDPMGVYVWLOGVHICHS---YNSIMWGQIIPQDWMDFQSDHCEAKFH 601
QY 494 AAARATGCPRIYVSDKPGNNHFDLKLKLVLPDGSVLRALQPERPRLDLPDPAARDKSL 553
Db 602 AGSRAICGSPRYVSDSVSGSHDEPLIKLVFPDGTVPKCIYFPPLPRDCLFRNPLFDOKTV 661
QY 554 LKTMNINKSGVYVGNOCGAGMCKIEKTRIHDTSPGTLTASYCASDVDLITQVAGAEW 613
Db 662 LKTMNINKSGVYVGAFCNCGAGMDPKGKFKGPPRECYRAISCTVAYTEW-DOKKEADH 720
QY 614 LG---DTIVAYRSGEVIRLPKGVSP--VTIKVLEFELHPCIOELAPS-ISFAIIGL 667
Db 721 MGAEEYVYV-LNQAVELHMTVPSEPLQTLTQPSFTELXNVPVEKLGSSNIRKPAIQL 779
QY 668 LDMFNTGAVEOVEIHNRAATKTALSVGRGRFVYSSQRPKCVVGAETDPNDYSET 727
Db 780 TNMFNSGGRIOELEI---IEKDVKYKVGKGRFLAYTSQSPKRFQLNGSDAFAFWLPDG 835
QY 728 GLT 730
Db 836 KLT 838

RESULT 13
Q93466 PRELIMINARY; PRT; 357 AA.
ID 093466:
AC 093466:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SEED IMBIBITION PROTEIN (FRAGMENT).
RN [1]
RP SEQUENCE FROM N.A.
RX Certences E.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; X95875; CAA65125.1; -.
FT NON_TER 1 1
FT NON_TER 357 357
SQ SEQUENCE 357 AA; 39465 MW; E8CD30399E1BF3EE CRC64;

Query Match 30.7%; Score 1228.5; DB 10; Length 357;
Best Local Similarity 68.0%; Pred. No. 1.8e-93;
Matches 247; Conservative 24; Mismatches 75; Indels 17; Gaps 6;

QY 172 QAYKAVKHNQOTFLHREKRLPSCLDPMFGCTWDATFYDTVTAEGVEGLKSLSQGTPPR 231
Db 4 QAYKAVKHNQOTFLHREKRLPSCLDPMFGCTWDATFYDTVTAEGVEGLKSLSQGTPPR 63
QY 232 FLITIDGWOQIENKAKATECLVQEGAGPATRLTG--IKENTKFOKKLONNEQMSGLKHL 289
Db 64 FLITVGNKRLKAYQSGSCVYQEGAGQ---LLGDMIRANRANKNNKGQNDQIPLKHL 120
```


